A user's guide to optimal transport

Luigi Ambrosio [*] Nicola Gigli [†]

Abstract

This text is an expanded version of the lectures given by the first author in the 2009 CIME summer school of Cetraro. It provides a quick and reasonably account of the classical theory of optimal mass transportation and of its more recent developments, including the metric theory of gradient flows, geometric and functional inequalities related to optimal transportation, the first and second order differential calculus in the Wasserstein space and the synthetic theory of metric measure spaces with Ricci curvature bounded from below.

Contents

[*] l.ambrosio@sns.it
[†] nicola.gigli@unice.fr

Introduction

The opportunity to write down these notes on Optimal Transport has been the CIME course in Cetraro given by the first author in 2009. Later on the second author joined to the project, and the initial set of notes has been enriched and made more detailed, in particular in connection with the differentiable structure of the Wasserstein space, the synthetic curvature bounds and their analytic implications. Some of the results presented here have not yet appeared in a book form, with the exception of [44].

It is clear that this subject is expanding so quickly that it is impossible to give an account of all developments of the theory in a few hours, or a few pages. A more modest approach is to give a quick mention of the many aspects of the theory, stimulating the reader's curiosity and leaving to more detailed treatises as [6] (mostly focused on the theory of gradient flows) and the monumental book [80] (for a -much - broader overview on optimal transport).

In Chapter 1 we introduce the optimal transport problem and its formulations in terms of transport maps and transport plans. Then we introduce basic tools of the theory, namely the duality formula, the c-monotonicity and discuss the problem of existence of optimal maps in the model case cost=distance^2.

In Chapter 2 we introduce the Wasserstein distance W_2 on the set $\mathscr{P}_2(X)$ of probability measures with finite quadratic moments and X is a generic Polish space. This distance naturally arises when considering the optimal transport problem with quadratic cost. The connections between geodesics in $\mathscr{P}_2(X)$ and geodesics in X and between the time evolution of Kantorovich potentials and the Hopf-Lax semigroup are discussed in detail. Also, when looking at geodesics in this space, and in particular when the underlying metric space X is a Riemannian manifold M, one is naturally lead to the so-called time-dependent optimal transport problem, where geodesics are singled out by an action minimization principle. This is the so-called Benamou-Brenier formula, which is the first step in the interpretation of $\mathscr{P}_2(M)$ as an infinite-dimensional Riemannian manifold, with W_2 as Riemannian distance. We then further exploit this viewpoint following Otto's seminal work [67].

In Chapter 3 we make a quite detailed introduction to the theory of gradient flows, borrowing almost all material from [6]. First we present the classical theory, for λ-convex functionals in Hilbert spaces. Then we present some equivalent formulations that involve only the distance, and therefore are applicable (at least in principle) to general metric space. They involve the derivative of the distance from a point (the (EVI) formulation) or the rate of dissipation of the energy (the (EDE) and (EDI) formulations). For all these formulations there is a corresponding discrete version of the gradient flow formulation given by the implicit Euler scheme. We will then show that there is convergence of the scheme to the continuous solution as the time discretization parameter tends to 0. The (EVI) formulation is the stronger one, in terms of uniqueness, contraction and regularizing effects. On the other hand this formulation depends on a compatibility condition between energy and distance; this condition is fulfilled in Non Positively Curved spaces in the sense of Alexandrov if the energy is convex along geodesics. Luckily enough, the compatibility condition holds even for some important model functionals in $\mathscr{P}_2(\mathbb{R}^n)$ (sum of the so-called internal, potential and interaction energies), even though the space is Positively Curved in the sense of Alexandrov.

In Chapter 4 we illustrate the power of optimal transportation techniques in the proof of some classical functional/geometric inequalities: the Brunn-Minkowski inequality, the isoperimetric inequality and the Sobolev inequality. Recent works in this area have also shown the possibility to prove by optimal transportation methods optimal effective versions of these inequalities: for instance we can quantify the closedness of E to a ball with the same volume in terms of the vicinity of the isoperimetric ratio of E to the optimal one.

Chapter 5 is devoted to the presentation of three recent variants of the optimal transport problem, which lead to different notions of Wasserstein distance: the first one deals with variational problems giving rise to branched transportation structures, with a 'Y shaped path' opposed to the 'V shaped one' typical of the mass splitting occurring in standard optimal transport problems. The second one involves modification in the action functional on curves arising in the Benamou-Brenier formula: this leads to many different optimal transportation distances, maybe more difficult to describe from the Lagrangian viepoint, but still with quite useful implications in evolution PDE's and functional inequalities. The last one deals with transportation distance between measures with unequal mass, a variant useful in the modeling problems with Dirichlet boundary conditions.

Chapter 6 deals with a more detailed analysis of the differentiable structure of $\mathscr{P}_2(\mathbb{R}^d)$: besides the analytic tangent space arising from the Benamou-Brenier formula, also the "geometric" tangent space, based on constant speed geodesics emanating from a given base point, is introduced. We also present Otto's viewpoint on the duality between Wasserstein space and Arnold's manifolds of measure-preserving diffeomorphisms. A large part of the chapter is also devoted to the second order differentiable properties, involving curvature. The notions of parallel transport along (sufficiently regular) geodesics and Levi-Civita connection in the Wasserstein space are discussed in detail.

Finally, Chapter 7 is devoted to an introduction to the synthetic notions of Ricci lower bounds for metric measure spaces introduced by Lott & Villani and Sturm in recent papers. This notion is based on suitable convexity properties of a dimension-dependent internal energy along Wasserstein geodesics. Synthetic Ricci bounds are completely consistent with the smooth Riemannian case and stable under measured-Gromov-Hausdorff limits. For this reason these bounds, and their analytic implications, are a useful tool in the description of measured-GH-limits of Riemannian manifolds.

Acknowledgement. Work partially supported by a MIUR PRIN2008 grant.

1 The optimal transport problem

1.1 Monge and Kantorovich formulations of the optimal transport problem

Given a Polish space (X, d) (i.e. a complete and separable metric space), we will denote by $\mathscr{P}(X)$ the set of Borel probability measures on X. By support $\mathrm{supp}(\mu)$ of a measure $\mu \in \mathscr{P}(X)$ we intend the smallest closed set on which μ is concentrated.

If X, Y are two Polish spaces, $T : X \to Y$ is a Borel map, and $\mu \in \mathscr{P}(X)$ a measure, the measure $T_\# \mu \in \mathscr{P}(Y)$, called the *push forward of μ through T* is defined by

$$T_\# \mu(E) = \mu(T^{-1}(E)), \qquad \forall E \subset Y, \text{ Borel.}$$

The push forward is characterized by the fact that

$$\int f dT_\# \mu = \int f \circ T d\mu,$$

for every Borel function $f : Y \to \mathbb{R} \cup \{\pm\infty\}$, where the above identity has to be understood in the following sense: one of the integrals exists (possibly attaining the value $\pm\infty$) if and only if the other one exists, and in this case the values are equal.

Now fix a Borel *cost function* $c : X \times Y \to \mathbb{R} \cup \{+\infty\}$. The Monge version of the transport problem is the following:

Problem 1.1 (Monge's optimal transport problem) *Let $\mu \in \mathscr{P}(X)$, $\nu \in \mathscr{P}(Y)$. Minimize*

$$T \mapsto \int_X c\big(x, T(x)\big)\, d\mu(x)$$

among all transport maps *T from μ to ν, i.e. all maps T such that $T_\# \mu = \nu$.* ■

Regardless of the choice of the cost function c, Monge's problem can be ill-posed because:

- no admissible T exists (for instance if μ is a Dirac delta and ν is not).
- the constraint $T_\# \mu = \nu$ is not weakly sequentially closed, w.r.t. any reasonable weak topology.

As an example of the second phenomenon, one can consider the sequence $f_n(x) := f(nx)$, where $f : \mathbb{R} \to \mathbb{R}$ is 1-periodic and equal to 1 on $[0, 1/2)$ and to -1 on $[1/2, 1)$, and the measures $\mu := \mathcal{L}|_{[0,1]}$ and $\nu := (\delta_{-1} + \delta_1)/2$. It is immediate to check that $(f_n)_\# \mu = \nu$ for every $n \in \mathbb{N}$, and yet (f_n) weakly converges to the null function $f \equiv 0$ which satisfies $f_\# \mu = \delta_0 \neq \nu$.

A way to overcome these difficulties is due to Kantorovich, who proposed the following way to relax the problem:

Problem 1.2 (Kantorovich's formulation of optimal transportation) *We minimize*

$$\boldsymbol{\gamma} \mapsto \int_{X \times Y} c(x, y)\, d\boldsymbol{\gamma}(x, y)$$

in the set $\mathcal{Adm}(\mu, \nu)$ of all transport plans *$\boldsymbol{\gamma} \in \mathscr{P}(X \times Y)$ from μ to ν, i.e. the set of Borel Probability measures on $X \times Y$ such that*

$$\boldsymbol{\gamma}(A \times Y) = \mu(A) \quad \forall A \in \mathscr{B}(X), \qquad \boldsymbol{\gamma}(X \times B) = \nu(B) \quad \forall B \in \mathscr{B}(Y).$$

Equivalently: $\pi^X_\# \boldsymbol{\gamma} = \mu$, $\pi^Y_\# \boldsymbol{\gamma} = \nu$, where π^X, π^Y are the natural projections from $X \times Y$ onto X and Y respectively. ■

Transport plans can be thought of as "multivalued" transport maps: $\boldsymbol{\gamma} = \int \boldsymbol{\gamma}_x \, d\mu(x)$, with $\boldsymbol{\gamma}_x \in \mathscr{P}(\{x\} \times Y)$. Another way to look at transport plans is to observe that for $\boldsymbol{\gamma} \in \mathcal{A}dm(\mu, \nu)$, the value of $\boldsymbol{\gamma}(A \times B)$ is the amount of mass initially in A which is sent into the set B.

There are several advantages in the Kantorovich formulation of the transport problem:

- $\mathcal{A}dm(\mu, \nu)$ is always not empty (it contains $\mu \times \nu$),
- the set $\mathcal{A}dm(\mu, \nu)$ is convex and compact w.r.t. the narrow topology in $\mathscr{P}(X \times Y)$ (see below for the definition of narrow topology and Theorem 1.5), and $\boldsymbol{\gamma} \mapsto \int c \, d\boldsymbol{\gamma}$ is linear,
- minima always exist under mild assumptions on c (Theorem 1.5),
- transport plans "include" transport maps, since $T_{\#}\mu = \nu$ implies that $\boldsymbol{\gamma} := (Id \times T)_{\#}\mu$ belongs to $\mathcal{A}dm(\mu, \nu)$.

In order to prove existence of minimizers of Kantorovich's problem we recall some basic notions concerning analysis over a Polish space. We say that a sequence $(\mu_n) \subset \mathscr{P}(X)$ *narrowly* converges to μ provided

$$\int \varphi \, d\mu_n \quad \mapsto \quad \int \varphi \, d\mu, \qquad \forall \varphi \in C_b(X),$$

$C_b(X)$ being the space of continuous and bounded functions on X. It can be shown that the topology of narrow convergence is metrizable. A set $\mathcal{K} \subset \mathscr{P}(X)$ is called *tight* provided for every $\varepsilon > 0$ there exists a compact set $K_\varepsilon \subset X$ such that

$$\mu(X \setminus K_\varepsilon) \leq \varepsilon, \qquad \forall \mu \in \mathcal{K}.$$

It holds the following important result.

Theorem 1.3 (Prokhorov) *Let (X, d) be a Polish space. Then a family $\mathcal{K} \subset \mathscr{P}(X)$ is relatively compact w.r.t. the narrow topology if and only if it is tight.*

Notice that if $\mathcal{K}$ contains only one measure, one recovers Ulam's theorem: any Borel probability measure on a Polish space is concentrated on a σ-compact set.

Remark 1.4 The inequality

$$\boldsymbol{\gamma}(X \times Y \setminus K_1 \times K_2) \leq \mu(X \setminus K_1) + \nu(Y \setminus K_2), \tag{1.1}$$

valid for any $\boldsymbol{\gamma} \in \mathcal{A}dm(\mu, \nu)$, shows that if $\mathcal{K}_1 \subset \mathscr{P}(X)$ and $\mathcal{K}_2 \subset \mathscr{P}(Y)$ are tight, then so is the set

$$\left\{ \boldsymbol{\gamma} \in \mathscr{P}(X \times Y) \ : \ \pi^X_{\#}\boldsymbol{\gamma} \in \mathcal{K}_1, \ \pi^Y_{\#}\boldsymbol{\gamma} \in \mathcal{K}_2 \right\}$$

■

Existence of minimizers for Kantorovich's formulation of the transport problem now comes from a standard lower-semicontinuity and compactness argument:

Theorem 1.5 *Assume that c is lower semicontinuous and bounded from below. Then there exists a minimizer for Problem 1.2.*

Proof
Compactness Remark 1.4 and Ulam's theorem show that the set $\mathcal{A}dm(\mu, \nu)$ is tight in $\mathscr{P}(X \times Y)$, and hence relatively compact by Prokhorov theorem.

To get the narrow compactness, pick a sequence $(\boldsymbol{\gamma}_n) \subset \mathcal{Adm}(\mu,\nu)$ and assume that $\boldsymbol{\gamma}_n \to \boldsymbol{\gamma}$ narrowly: we want to prove that $\boldsymbol{\gamma} \in \mathcal{Adm}(\mu,\nu)$ as well. Let φ be any function in $C_b(X)$ and notice that $(x,y) \mapsto \varphi(x)$ is continuous and bounded in $X \times Y$, hence we have

$$\int \varphi \, d\pi^X_\# \boldsymbol{\gamma} = \int \varphi(x) \, d\boldsymbol{\gamma}(x,y) = \lim_{n\to\infty} \int \varphi(x) \, d\boldsymbol{\gamma}_n(x,y) = \lim_{n\to\infty} \int \varphi \, d\pi^X_\# \boldsymbol{\gamma}_n = \int \varphi \, d\mu,$$

so that by the arbitrariness of $\varphi \in C_b(X)$ we get $\pi^X_\# \boldsymbol{\gamma} = \mu$. Similarly we can prove $\pi^Y_\# \boldsymbol{\gamma} = \nu$, which gives $\boldsymbol{\gamma} \in \mathcal{Adm}(\mu,\nu)$ as desired.
Lower semicontinuity. We claim that the functional $\boldsymbol{\gamma} \mapsto \int c \, d\boldsymbol{\gamma}$ is l.s.c. with respect to narrow convergence. This is true because our assumptions on c guarantee that there exists an increasing sequence of functions $c_n : X \times Y \to \mathbb{R}$ continuous an bounded such that $c(x,y) = \sup_n c_n(x,y)$, so that by monotone convergence it holds

$$\int c \, d\boldsymbol{\gamma} = \sup_n \int c_n \, d\boldsymbol{\gamma}.$$

Since by construction $\boldsymbol{\gamma} \mapsto \int c_n \, d\boldsymbol{\gamma}$ is narrowly continuous, the proof is complete. □

We will denote by $\mathcal{Opt}(\mu,\nu)$ the set of *optimal plans* from μ to ν for the Kantorovich formulation of the transport problem, i.e. the set of minimizers of Problem 1.2. More generally, we will say that a plan is optimal, if it is optimal between its own marginals. Observe that with the notation $\mathcal{Opt}(\mu,\nu)$ we are losing the reference to the cost function c, which of course affects the set itself, but the context will always clarify the cost we are referring to.

Once existence of optimal plans is proved, a number of natural questions arise:

- are optimal plans unique?
- is there a simple way to check whether a given plan is optimal or not?
- do optimal plans have any natural regularity property? In particular, are they induced by maps?
- how far is the minimum of Problem 1.2 from the infimum of Problem 1.1?

This latter question is important to understand whether we can really consider Problem 1.2 the relaxation of Problem 1.1 or not. It is possible to prove that if c is continuous and μ is non atomic, then

$$\inf \text{(Monge)} = \min \text{(Kantorovich)}, \tag{1.2}$$

so that transporting with plans can't be strictly cheaper than transporting with maps. We won't detail the proof of this fact.

1.2 Necessary and sufficient optimality conditions

To understand the structure of optimal plans, probably the best thing to do is to start with an example.

Let $X = Y = \mathbb{R}^d$ and $c(x,y) := |x-y|^2/2$. Also, assume that $\mu, \nu \in \mathscr{P}(\mathbb{R}^d)$ are supported on finite sets. Then it is immediate to verify that a plan $\boldsymbol{\gamma} \in \mathcal{Adm}(\mu,\nu)$ is optimal if and only if it holds

$$\sum_{i=1}^N \frac{|x_i - y_i|^2}{2} \le \sum_{i=1}^N \frac{|x_i - y_{\sigma(i)}|^2}{2},$$

for any $N \in \mathbb{N}$, $(x_i, y_i) \in \operatorname{supp}(\boldsymbol{\gamma})$ and σ permutation of the set $\{1,\ldots,N\}$. Expanding the squares we get

$$\sum_{i=1}^N \langle x_i, y_i \rangle \ge \sum_{i=1}^N \langle x_i, y_{\sigma(i)} \rangle,$$

which by definition means that the support of γ is cyclically monotone. Let us recall the following theorem:

Theorem 1.6 (Rockafellar) *A set $\Gamma \subset \mathbb{R}^d \times \mathbb{R}^d$ is cyclically monotone if and only if there exists a convex and lower semicontinuous function $\varphi : \mathbb{R}^d \to \mathbb{R} \cup \{+\infty\}$ such that Γ is included in the graph of the subdifferential of φ.*

We skip the proof of this theorem, because later on we will prove a much more general version. What we want to point out here is that under the above assumptions on μ and ν we have that the following three things are equivalent:

- $\gamma \in \mathcal{Adm}(\mu, \nu)$ is optimal,

- $\mathrm{supp}(\gamma)$ is cyclically monotone,

- there exists a convex and lower semicontinuous function φ such that γ is concentrated on the graph of the subdifferential of φ.

The good news is that the equivalence between these three statements holds in a much more general context (more general underlying spaces, cost functions, measures). Key concepts that are needed in the analysis, are the generalizations of the concepts of cyclical monotonicity, convexity and subdifferential which fit with a general cost function c.

The definitions below make sense for a general Borel and real valued cost.

Definition 1.7 (c-cyclical monotonicity) *We say that $\Gamma \subset X \times Y$ is* c-cyclically monotone *if $(x_i, y_i) \in \Gamma$, $1 \le i \le N$, implies*

$$\sum_{i=1}^{N} c(x_i, y_i) \le \sum_{i=1}^{N} c(x_i, y_{\sigma(i)}) \qquad \textit{for all permutations } \sigma \textit{ of } \{1, \dots, N\}.$$

Definition 1.8 (c-transforms) *Let $\psi : Y \to \mathbb{R} \cup \{\pm\infty\}$ be any function. Its c_+-transform $\psi^{c+} : X \to \mathbb{R} \cup \{-\infty\}$ is defined as*

$$\psi^{c+}(x) := \inf_{y \in Y} c(x, y) - \psi(y).$$

Similarly, given $\varphi : X \to \mathbb{R} \cup \{\pm\infty\}$, its c_+-transform is the function $\varphi^{c+} : Y \to \mathbb{R} \cup \{\pm\infty\}$ defined by

$$\varphi^{c+}(y) := \inf_{x \in X} c(x, y) - \varphi(x).$$

The c_--transform $\psi^{c-} : X \to \mathbb{R} \cup \{+\infty\}$ of a function ψ on Y is given by

$$\psi^{c-}(x) := \sup_{y \in Y} -c(x, y) - \psi(y),$$

and analogously for c_--transforms of functions φ on X.

Definition 1.9 (c-concavity and c-convexity) *We say that $\varphi : X \to \mathbb{R} \cup \{-\infty\}$ is c-concave if there exists $\psi : Y \to \mathbb{R} \cup \{-\infty\}$ such that $\varphi = \psi^{c+}$. Similarly, $\psi : Y \to \mathbb{R} \cup \{-\infty\}$ is c-concave if there exists $\varphi : Y \to \mathbb{R} \cup \{-\infty\}$ such that $\psi = \varphi^{c+}$.*

Symmetrically, $\varphi : X \to \mathbb{R} \cup \{+\infty\}$ is c-convex if there exists $\psi : Y \to \mathbb{R} \cup \{+\infty\}$ such that $\varphi = \psi^{c-}$, and $\psi : Y \to \mathbb{R} \cup \{+\infty\}$ is c-convex if there exists $\varphi : Y \to \mathbb{R} \cup \{+\infty\}$ such that $\psi = \varphi^{c-}$.

Observe that $\varphi : X \to \mathbb{R} \cup \{-\infty\}$ is c-concave if and only if $\varphi^{c_+ c_+} = \varphi$. This is a consequence of the fact that for any function $\psi : Y \to \mathbb{R} \cup \{\pm\infty\}$ it holds $\psi^{c_+} = \psi^{c_+ c_+ c_+}$, indeed

$$\psi^{c_+ c_+ c_+}(x) = \inf_{\tilde{y} \in Y} \sup_{\tilde{x} \in X} \inf_{y \in Y} c(x, \tilde{y}) - c(\tilde{x}, \tilde{y}) + c(\tilde{x}, y) - \psi(y),$$

and choosing $\tilde{x} = x$ we get $\psi^{c_+ c_+ c_+} \geq \psi^{c_+}$, while choosing $y = \tilde{y}$ we get $\psi^{c_+ c_+ c_+} \leq \psi^{c_+}$. Similarly for functions on Y and for the c-convexity.

Definition 1.10 (c-superdifferential and c-subdifferential) *Let $\varphi : X \to \mathbb{R} \cup \{-\infty\}$ be a c-concave function. The c-superdifferential $\partial^{c_+}\varphi \subset X \times Y$ is defined as*

$$\partial^{c_+}\varphi := \Big\{ (x, y) \in X \times Y \ : \ \varphi(x) + \varphi^{c_+}(y) = c(x, y) \Big\}.$$

The c-superdifferential $\partial^{c_+}\varphi(x)$ at $x \in X$ is the set of $y \in Y$ such that $(x, y) \in \partial^{c_+}\varphi$. A symmetric definition is given for c-concave functions $\psi : Y \to \mathbb{R} \cup \{-\infty\}$.

The definition of c-subdifferential ∂^{c_-} of a c-convex function $\varphi : X \to \{+\infty\}$ is analogous:

$$\partial^{c_-}\varphi := \Big\{ (x, y) \in X \times Y \ : \ \varphi(x) + \varphi^{c_-}(y) = -c(x, y) \Big\}.$$

Analogous definitions hold for c-concave and c-convex functions on Y.

Remark 1.11 (The base case: $c(x, y) = -\langle x, y \rangle$) Let $X = Y = \mathbb{R}^d$ and $c(x, y) = -\langle x, y \rangle$. Then a direct application of the definitions show that:

- a set is c-cyclically monotone if and only if it is cyclically monotone
- a function is c-convex (resp. c-concave) if and only if it is convex and lower semicontinuous (resp. concave and upper semicontinuous),
- the c-subdifferential of the c-convex (resp. c-superdifferential of the c-concave) function is the classical subdifferential (resp. superdifferential),
- the c_- transform is the Legendre transform.

Thus in this situation these new definitions become the classical basic definitions of convex analysis. ■

Remark 1.12 (For most applications c-concavity is sufficient) There are several trivial relations between c-convexity, c-concavity and related notions. For instance, φ is c-concave if and only if $-\varphi$ is c-convex, $-\varphi^{c_+} = (-\varphi)^{c_-}$ and $\partial^{c_+}\varphi = \partial^{c_-}(-\varphi)$. Therefore, roughly said, every statement concerning c-concave functions can be restated in a statement for c-convex ones. Thus, choosing to work with c-concave or c-convex functions is actually a matter of taste.

Our choice is to work with c-concave functions. Thus all the statements from now on will deal only with these functions. There is only one, important, part of the theory where the distinction between c-concavity and c-convexity is useful: in the study of geodesics in the Wasserstein space (see Section 2.2, and in particular Theorem 2.18 and its consequence Corollary 2.24).

We also point out that the notation used here is different from the one in [80], where a less symmetric notion (but better fitting the study of geodesics) of c-concavity and c-convexity has been preferred. ■

An equivalent characterization of the c-superdifferential is the following: $y \in \partial^{c_+}\varphi(x)$ if and only if it holds

$$\begin{aligned} &\varphi(x) = c(x, y) - \varphi^{c_+}(y), \\ &\varphi(z) \leq c(z, y) - \varphi^{c_+}(y), \qquad \forall z \in X, \end{aligned}$$

or equivalently if

$$\varphi(x) - c(x, y) \geq \varphi(z) - c(z, y), \qquad \forall z \in X. \tag{1.3}$$

A direct consequence of the definition is that the c-superdifferential of a c-concave function is always a c-cyclically monotone set, indeed if $(x_i, y_i) \in \partial^{c+}\varphi$ it holds

$$\sum_i c(x_i, y_i) = \sum_i \varphi(x_i) + \varphi^c(y_i) = \sum_i \varphi(x_i) + \varphi^c(y_{\sigma(i)}) \leq \sum_i c(x_i, y_{\sigma(i)}),$$

for any permutation σ of the indexes.

What is important to know is that actually under mild assumptions on c, *every* c-cyclically monotone set can be obtained as the c-superdifferential of a c-concave function. This result is part of the following important theorem:

Theorem 1.13 (Fundamental theorem of optimal transport) *Assume that $c : X \times Y \to \mathbb{R}$ is continuous and bounded from below and let $\mu \in \mathscr{P}(X)$, $\nu \in \mathscr{P}(Y)$ be such that*

$$c(x, y) \leq a(x) + b(y), \tag{1.4}$$

for some $a \in L^1(\mu)$, $b \in L^1(\nu)$. Also, let $\boldsymbol{\gamma} \in \mathcal{A}dm(\mu, \nu)$. Then the following three are equivalent:

i) the plan $\boldsymbol{\gamma}$ is optimal,

ii) the set $\operatorname{supp}(\boldsymbol{\gamma})$ is c-cyclically monotone,

iii) there exists a c-concave function φ such that $\max\{\varphi, 0\} \in L^1(\mu)$ and $\operatorname{supp}(\boldsymbol{\gamma}) \subset \partial^{c+}\varphi$.

Proof Observe that the inequality (1.4) together with

$$\int c(x, y) d\tilde{\boldsymbol{\gamma}}(x, y) \leq \int a(x) + b(y) d\tilde{\boldsymbol{\gamma}}(x, y) = \int a(x) d\mu(x) + \int b(y) d\nu(y) < \infty, \quad \forall \tilde{\boldsymbol{\gamma}} \in \mathcal{A}dm(\mu, \nu)$$

implies that for any admissible plan $\tilde{\boldsymbol{\gamma}} \in \mathcal{A}dm(\mu, \nu)$ the function $\max\{c, 0\}$ is integrable. This, together with the bound from below on c gives that $c \in L^1(\tilde{\boldsymbol{\gamma}})$ for any admissible plan $\tilde{\boldsymbol{\gamma}}$.

$(\mathbf{i}) \Rightarrow (\mathbf{ii})$ We argue by contradiction: assume that the support of $\boldsymbol{\gamma}$ is not c-cyclically monotone. Thus we can find $N \in \mathbb{N}$, $\{(x_i, y_i)\}_{1 \leq i \leq N} \subset \operatorname{supp}(\boldsymbol{\gamma})$ and some permutation σ of $\{1, \ldots, N\}$ such that

$$\sum_{i=1}^N c(x_i, y_i) > \sum_{i=1}^N c(x_i, y_{\sigma(i)}).$$

By continuity we can find neighborhoods $U_i \ni x_i$, $V_i \ni y_i$ with

$$\sum_{i=1}^N c(u_i, v_{\sigma(i)}) - c(u_i, v_i) < 0 \qquad \forall (u_i, v_i) \in U_i \times V_i,\ 1 \leq i \leq N.$$

Our goal is to build a "variation" $\tilde{\boldsymbol{\gamma}} = \boldsymbol{\gamma} + \boldsymbol{\eta}$ of $\boldsymbol{\gamma}$ in such a way that minimality of $\boldsymbol{\gamma}$ is violated. To this aim, we need a *signed* measure $\boldsymbol{\eta}$ with:

(A) $\boldsymbol{\eta}^- \leq \boldsymbol{\gamma}$ (so that $\tilde{\boldsymbol{\gamma}}$ is nonnegative);

(B) null first and second marginal (so that $\tilde{\boldsymbol{\gamma}} \in \mathcal{A}dm(\mu, \nu)$);

(C) $\int c \, d\boldsymbol{\eta} < 0$ (so that $\boldsymbol{\gamma}$ is not optimal).

Let $\Omega := \Pi_{i=1}^N U_i \times V_i$ and $\mathbf{P} \in \mathscr{P}(\Omega)$ be defined as the product of the measures $\frac{1}{m_i}\boldsymbol{\gamma}|_{U_i \times V_i}$, where $m_i := \boldsymbol{\gamma}(U_i \times V_i)$. Denote by π^{U_i}, π^{V_i} the natural projections of Ω to U_i and V_i respectively and define

$$\boldsymbol{\eta} := \frac{\min_i m_i}{N} \sum_{i=i}^{N} (\pi^{U_i}, \pi^{V_{\sigma(i)}})_{\#}\mathbf{P} - (\pi^{U_i}, \pi^{V_{(i)}})_{\#}\mathbf{P}.$$

It is immediate to verify that $\boldsymbol{\eta}$ fulfills (A), (B), (C) above, so that the thesis is proven.

(**ii**) $\Rightarrow$ (**iii**) We need to prove that if $\Gamma \subset X \times Y$ is a c-cyclically monotone set, then there exists a c-concave function φ such that $\partial^c\varphi \supset \Gamma$ and $\max\{\varphi, 0\} \in L^1(\mu)$. Fix $(\overline{x}, \overline{y}) \in \Gamma$ and observe that, since we want φ to be c-concave with the c-superdifferential that contains Γ, for any choice of $(x_i, y_i) \in \Gamma$, $i = 1, \ldots, N$, we need to have

$$\begin{aligned}
\varphi(x) &\leq c(x, y_1) - \varphi^{c_+}(y_1) = c(x, y_1) - c(x_1, y_1) + \varphi(x_1) \\
&\leq \Big(c(x, y_1) - c(x_1, y_1)\Big) + c(x_1, y_2) - \varphi^{c_+}(y_2) \\
&= \Big(c(x, y_1) - c(x_1, y_1)\Big) + \Big(c(x_1, y_2) - c(x_2, y_2)\Big) + \varphi(x_2) \\
&\leq \cdots \\
&\leq \Big(c(x, y_1) - c(x_1, y_1)\Big) + \Big(c(x_1, y_2) - c(x_2, y_2)\Big) + \cdots + \Big(c(x_N, \overline{y}) - c(\overline{x}, \overline{y})\Big) + \varphi(\overline{x}).
\end{aligned}$$

It is therefore natural to define φ as the infimum of the above expression as $\{(x_i, y_i)\}_{i=1,\ldots,N}$ vary among all N-ples in Γ and N varies in $\mathbb{N}$. Also, since we are free to add a constant to φ, we can neglect the addendum $\varphi(\overline{x})$ and define:

$$\varphi(x) := \inf \Big(c(x, y_1) - c(x_1, y_1)\Big) + \Big(c(x_1, y_2) - c(x_2, y_2)\Big) + \cdots + \Big(c(x_N, \overline{y}) - c(\overline{x}, \overline{y})\Big),$$

the infimum being taken on $N \geq 1$ integer and $(x_i, y_i) \in \Gamma$, $i = 1, \ldots, N$. Choosing $N = 1$ and $(x_1, y_1) = (\overline{x}, \overline{y})$ we get $\varphi(\overline{x}) \leq 0$. Conversely, from the c-cyclical monotonicity of Γ we have $\varphi(\overline{x}) \geq 0$. Thus $\varphi(\overline{x}) = 0$.

Also, it is clear from the definition that φ is c-concave. Choosing again $N = 1$ and $(x_1, y_1) = (\overline{x}, \overline{y})$, using (1.3) we get

$$\varphi(x) \leq c(x, \overline{y}) - c(\overline{x}, \overline{y}) < a(x) + b(\overline{y}) - c(\overline{x}, \overline{y}),$$

which, together with the fact that $a \in L^1(\mu)$, yields $\max\{\varphi, 0\} \in L^1(\mu)$. Thus, we need only to prove that $\partial^{c_+}\varphi$ contains Γ. To this aim, choose $(\tilde{x}, \tilde{y}) \in \Gamma$, let $(x_1, y_1) = (\tilde{x}, \tilde{y})$ and observe that by definition of $\varphi(x)$ we have

$$\begin{aligned}
\varphi(x) &\leq c(x, \tilde{y}) - c(\tilde{x}, \tilde{y}) + \inf \Big(c(\tilde{x}, y_2) - c(x_2, y_2)\Big) + \cdots + \Big(c(x_N, \overline{y}) - c(\overline{x}, \overline{y})\Big) \\
&= c(x, \tilde{y}) - c(\tilde{x}, \tilde{y}) + \varphi(\tilde{x}).
\end{aligned}$$

By the characterization (1.3), this inequality shows that $(\tilde{x}, \tilde{y}) \in \partial^{c_+}\varphi$, as desired.

(**iii**) $\Rightarrow$ (**i**). Let $\tilde{\boldsymbol{\gamma}} \in \mathcal{Adm}(\mu, \nu)$ be any transport plan. We need to prove that $\int c\,d\boldsymbol{\gamma} \leq \int c\,d\tilde{\boldsymbol{\gamma}}$. Recall that we have

$$\begin{aligned}
\varphi(x) + \varphi^{c_+}(y) &= c(x, y), \qquad \forall (x, y) \in \operatorname{supp}(\boldsymbol{\gamma}) \\
\varphi(x) + \varphi^{c_+}(y) &\leq c(x, y), \qquad \forall x \in X,\ y \in Y,
\end{aligned}$$

and therefore

$$\int c(x,y)d\boldsymbol{\gamma}(x,y) = \int \varphi(x) + \varphi^{c+}(y)d\boldsymbol{\gamma}(x,y) = \int \varphi(x)d\mu(x) + \int \varphi^{c+}(y)d\nu(y)$$
$$= \int \varphi(x) + \varphi^{c+}(y)d\tilde{\boldsymbol{\gamma}}(x,y) \le \int c(x,y)d\tilde{\boldsymbol{\gamma}}(x,y).$$

□

Remark 1.14 Condition (1.4) is natural in some, but not all, problems. For instance problems with constraints or in Wiener spaces (infinite-dimensional Gaussian spaces) include $+\infty$-valued costs, with a "large" set of points where the cost is not finite. We won't discuss these topics. ■

An important consequence of the previous theorem is that being optimal is a property that depends only on the support of the plan $\boldsymbol{\gamma}$, and not on how the mass is distributed in the support itself: if $\boldsymbol{\gamma}$ is an optimal plan (between its own marginals) and $\tilde{\boldsymbol{\gamma}}$ is such that $\mathrm{supp}(\tilde{\boldsymbol{\gamma}}) \subset \mathrm{supp}(\boldsymbol{\gamma})$, then $\tilde{\boldsymbol{\gamma}}$ is optimal as well (between its own marginals, of course). We will see in Proposition 2.5 that one of the important consequences of this fact is the *stability of optimality*.

Analogous arguments works for maps. Indeed assume that $T : X \to Y$ is a map such that $T(x) \in \partial^{c+}\varphi(x)$ for some c-concave function φ for all x. Then, for every $\mu \in \mathscr{P}(X)$ such that condition (1.4) is satisfied for $\nu = T_\#\mu$, the map T is optimal between μ and $T_\#\mu$. Therefore it makes sense to say that T is an optimal map, without explicit mention to the reference measures.

Remark 1.15 From Theorem 1.13 we know that given $\mu \in \mathscr{P}(X)$, $\nu \in \mathscr{P}(Y)$ satisfying the assumption of the theorem, for every optimal plan $\boldsymbol{\gamma}$ there exists a c-concave function φ such that $\mathrm{supp}(\boldsymbol{\gamma}) \subset \partial^{c+}\varphi$. Actually, a stronger statement holds, namely: if $\mathrm{supp}(\boldsymbol{\gamma}) \subset \partial^{c+}\varphi$ for some optimal $\boldsymbol{\gamma}$, then $\mathrm{supp}(\boldsymbol{\gamma}') \subset \partial^{c+}\varphi$ for *every* optimal plan $\boldsymbol{\gamma}'$. Indeed arguing as in the proof of 1.13 one can see that $\max\{\varphi, 0\} \in L^1(\mu)$ implies $\max\{\varphi^{c+}, 0\} \in L^1(\nu)$ and thus it holds

$$\int \varphi d\mu + \int \varphi^{c+} d\nu = \int \varphi(x) + \varphi^{c+}(y)d\boldsymbol{\gamma}'(x,y) \le \int c(x,y)d\boldsymbol{\gamma}'(x,y) = \int c(x,y)d\boldsymbol{\gamma}(x,y)$$
$$\big(\mathrm{supp}(\boldsymbol{\gamma}) \subset \partial^{c+}\varphi\big) \quad = \int \varphi(x) + \varphi^{c+}(y)d\boldsymbol{\gamma}(x,y) = \int \varphi d\mu + \int \varphi^{c+} d\nu.$$

Thus the inequality must be an equality, which is true if and only if for $\boldsymbol{\gamma}'$-a.e. (x,y) it holds $(x,y) \in \partial^{c+}\varphi$, hence, by the continuity of c, we conclude $\mathrm{supp}(\boldsymbol{\gamma}') \subset \partial^{c+}\varphi$. ■

1.3 The dual problem

The transport problem in the Kantorovich formulation is the problem of minimizing the linear functional $\boldsymbol{\gamma} \mapsto \int c d\boldsymbol{\gamma}$ with the affine constraints $\pi^X_\#\boldsymbol{\gamma} = \mu$, $\pi^Y_\#\boldsymbol{\gamma} = \nu$ and $\boldsymbol{\gamma} \ge 0$. It is well known that problems of this kind admit a natural dual problem, where we maximize a linear functional with affine constraints. In our case the dual problem is:

Problem 1.16 (Dual problem) *Let $\mu \in \mathscr{P}(X)$, $\nu \in \mathscr{P}(Y)$. Maximize the value of*

$$\int \varphi(x)d\mu(x) + \int \psi(y)d\nu(y),$$

among all functions $\varphi \in L^1(\mu)$, $\psi \in L^1(\nu)$ such that

$$\varphi(x) + \psi(y) \le c(x,y), \qquad \forall x \in X,\ y \in Y. \tag{1.5}$$

■

The relation between the transport problem and the dual one consists in the fact that

$$\inf_{\boldsymbol{\gamma}\in \mathcal{A}dm(\mu,\nu)} \int c(x,y)d\boldsymbol{\gamma}(x,y) = \sup_{\varphi,\psi} \int \varphi(x)d\mu(x) + \int \psi(y)d\nu(y),$$

where the supremum is taken among all φ, ψ as in the definition of the problem.

Although the fact that equality holds is an easy consequence of Theorem 1.13 of the previous section (taking $\psi = \varphi^{c+}$, as we will see), we prefer to start with an heuristic argument which shows "why" duality works. The calculations we are going to do are very common in linear programming and are based on the *min-max principle*. Observe how the constraint $\boldsymbol{\gamma} \in \mathcal{A}dm(\mu,\nu)$ "becomes" the functional to maximize in the dual problem and the functional to minimize $\int c d\boldsymbol{\gamma}$ "becomes" the constraint in the dual problem.

Start observing that

$$\inf_{\boldsymbol{\gamma}\in \mathcal{A}dm(\mu,\nu)} \int c(x,y)d\boldsymbol{\gamma}(x,y) = \inf_{\boldsymbol{\gamma}\in \mathcal{M}_+(X\times Y)} \int c(x,y)d\boldsymbol{\gamma} + \chi(\boldsymbol{\gamma}), \tag{1.6}$$

where $\chi(\boldsymbol{\gamma})$ is equal to 0 if $\boldsymbol{\gamma} \in \mathcal{A}dm(\mu,\nu)$ and $+\infty$ if $\boldsymbol{\gamma} \notin \mathcal{A}dm(\mu,\nu)$, and $\mathcal{M}_+(X\times Y)$ is the set of non negative Borel measures on $X \times Y$. We claim that the function χ may be written as

$$\chi(\boldsymbol{\gamma}) = \sup_{\varphi,\psi} \Big\{ \int \varphi(x)d\mu(x) + \int \psi(y)d\nu(y) - \int \varphi(x) + \psi(y)d\boldsymbol{\gamma}(x,y) \Big\},$$

where the supremum is taken among all $(\varphi,\psi) \in C_b(X) \times C_b(Y)$. Indeed, if $\boldsymbol{\gamma} \in \mathcal{A}dm(\mu,\nu)$ then $\chi(\boldsymbol{\gamma}) = 0$, while if $\boldsymbol{\gamma} \notin \mathcal{A}dm(\mu,\nu)$ we can find $(\varphi,\psi) \in C_b(X)\times C_b(Y)$ such that the value between the brackets is different from 0, thus by multiplying (φ,ψ) by appropriate real numbers we have that the supremum is $+\infty$. Thus from (1.6) we have

$$\begin{aligned}
&\inf_{\boldsymbol{\gamma}\in \mathcal{A}dm(\mu,\nu)} \int c(x,y)d\boldsymbol{\gamma}(x,y) \\
&= \inf_{\boldsymbol{\gamma}\in \mathcal{M}_+(X\times Y)} \sup_{\varphi,\psi} \left\{ \int c(x,y)d\boldsymbol{\gamma}(x,y) + \int \varphi(x)d\mu(x) + \int \psi(y)d\nu(y) - \int \varphi(x) + \psi(y)d\boldsymbol{\gamma}(x,y) \right\}.
\end{aligned}$$

Call the expression between brackets $F(\boldsymbol{\gamma},\varphi,\psi)$. Since $\boldsymbol{\gamma} \mapsto F(\boldsymbol{\gamma},\varphi,\psi)$ is convex (actually linear) and $(\varphi,\psi) \mapsto F(\boldsymbol{\gamma},\varphi,\psi)$ is concave (actually linear), the min-max principle holds and we have

$$\inf_{\boldsymbol{\gamma}\in \mathcal{A}dm(\mu,\nu)} \sup_{\varphi,\psi} F(\boldsymbol{\gamma},\varphi,\psi) = \sup_{\varphi,\psi} \inf_{\boldsymbol{\gamma}\in \mathcal{M}_+(X\times Y)} F(\boldsymbol{\gamma},\varphi,\psi).$$

Thus we have

$$\begin{aligned}
&\inf_{\boldsymbol{\gamma}\in \mathcal{A}dm(\mu,\nu)} \int c(x,y)d\boldsymbol{\gamma}(x,y) \\
&= \sup_{\varphi,\psi} \inf_{\boldsymbol{\gamma}\in \mathcal{M}_+(X\times Y)} \left\{ \int c(x,y)d\boldsymbol{\gamma}(x,y) + \int \varphi(x)d\mu(x) + \int \psi(y)d\nu(y) - \int \varphi(x) + \psi(y)d\boldsymbol{\gamma}(x,y) \right\} \\
&= \sup_{\varphi,\psi} \left\{ \int \varphi(x)d\mu(x) + \int \psi(y)d\nu(y) + \inf_{\boldsymbol{\gamma}\in \mathcal{M}_+(X\times Y)} \left[\int c(x,y) - \varphi(x) - \psi(y)d\boldsymbol{\gamma}(x,y) \right] \right\}.
\end{aligned}$$

Now observe the quantity

$$\inf_{\boldsymbol{\gamma}\in \mathcal{M}_+(X\times Y)} \left[\int c(x,y) - \varphi(x) - \psi(y)d\boldsymbol{\gamma}(x,y) \right].$$

If $\varphi(x) + \psi(y) \leq c(x,y)$ for any (x,y), then the integrand is non-negative and the infimum is 0 (achieved when $\boldsymbol{\gamma}$ is the null-measure). Conversely, if $\varphi(x) + \psi(y) > c(x,y)$ for some $(x,y) \in X \times Y$, then choose $\boldsymbol{\gamma} := n\delta_{(x,y)}$ with n large to get that the infimum is $-\infty$.

Thus, we proved that

$$\inf_{\boldsymbol{\gamma} \in \mathcal{Adm}(\mu,\nu)} \int c(x,y) d\boldsymbol{\gamma}(x,y) = \sup_{\varphi,\psi} \int \varphi(x) d\mu(x) + \int \psi(y) d\nu(y),$$

where the supremum is taken among continuous and bounded functions (φ, ψ) satisfying (1.5).

We now give the rigorous statement and a proof independent of the min-max principle.

Theorem 1.17 (Duality) *Let $\mu \in \mathscr{P}(X)$, $\nu \in \mathscr{P}(Y)$ and $c : X \times Y \to \mathbb{R}$ a continuous and bounded from below cost function. Assume that* (1.4) *holds. Then the minimum of the Kantorovich problem 1.2 is equal to the supremum of the dual problem 1.16.*
Furthermore, the supremum of the dual problem is attained, and the maximizing couple (φ, ψ) is of the form (φ, φ^{c+}) for some c-concave function φ.

Proof Let $\boldsymbol{\gamma} \in \mathcal{Adm}(\mu,\nu)$ and observe that for any couple of functions $\varphi \in L^1(\mu)$ and $\psi \in L^1(\nu)$ satisfying (1.5) it holds

$$\int c(x,y) d\boldsymbol{\gamma}(x,y) \geq \int \varphi(x) + \psi(y) d\boldsymbol{\gamma}(x,y) = \int \varphi(x) d\mu(x) + \int \psi(y) d\nu(y).$$

This shows that the minimum of the Kantorovich problem is $\geq$ than the supremum of the dual problem.

To prove the converse inequality pick $\boldsymbol{\gamma} \in \mathcal{Opt}(\mu,\nu)$ and use Theorem 1.13 to find a c-concave function φ such that $\mathrm{supp}(\boldsymbol{\gamma}) \subset \partial^{c+}\varphi$, $\max\{\varphi, 0\} \in L^1(\mu)$ and $\max\{\varphi^{c+}, 0\} \in L^1(\nu)$. Then, as in the proof of $(iii) \Rightarrow (i)$ of Theorem 1.13, we have

$$\int c(x,y)\, d\boldsymbol{\gamma}(x,y) = \int \varphi(x) + \varphi^{c+}(y)\, d\boldsymbol{\gamma}(x,y) = \int \varphi(x)\, d\mu(x) + \int \varphi^{c+}(y)\, d\nu(y),$$

and $\int c\, d\boldsymbol{\gamma} \in \mathbb{R}$. Thus $\varphi \in L^1(\mu)$ and $\varphi^{c+} \in L^1(\nu)$, which shows that (φ, φ^{c+}) is an admissible couple in the dual problem and gives the thesis. □

Remark 1.18 Notice that a statement stronger than the one of Remark 1.15 holds, namely: under the assumptions of Theorems 1.13 and 1.17, for any c-concave couple of functions (φ, φ^{c+}) maximizing the dual problem and any optimal plan $\boldsymbol{\gamma}$ it holds

$$\mathrm{supp}(\boldsymbol{\gamma}) \subset \partial^{c+}\varphi.$$

Indeed we already know that for some c-concave φ we have $\varphi \in L^1(\mu)$, $\varphi^{c+} \in L^1(\nu)$ and

$$\mathrm{supp}(\boldsymbol{\gamma}) \subset \partial^{c+}\varphi,$$

for any optimal $\boldsymbol{\gamma}$. Now pick another maximizing couple $(\tilde{\varphi}, \tilde{\psi})$ for the dual problem 1.16 and notice that $\tilde{\varphi}(x) + \tilde{\psi}(y) \leq c(x,y)$ for any x, y implies $\tilde{\psi} \leq \tilde{\varphi}^{c+}$, and therefore $(\tilde{\varphi}, \tilde{\varphi}^{c+})$ is a maximizing couple as well. The fact that $\tilde{\varphi}^{c+} \in L^1(\nu)$ follows as in the proof of Theorem 1.17. Conclude noticing that for any optimal plan $\boldsymbol{\gamma}$ it holds

$$\begin{aligned}\int \tilde{\varphi} d\mu + \int \tilde{\varphi}^{c+} d\nu &= \int \varphi d\mu + \int \varphi^{c+} d\nu = \int \varphi(x) + \varphi^{c+}(y) d\boldsymbol{\gamma}(x,y) \\ &= \int c(x,y) d\boldsymbol{\gamma} \geq \int \tilde{\varphi} d\mu + \int \tilde{\varphi}^{c+} d\nu,\end{aligned}$$

so that the inequality must be an equality. ■

Definition 1.19 (Kantorovich potential) *A c-concave function φ such that (φ, φ^{c_+}) is a maximizing pair for the dual problem 1.16 is called a c-concave Kantorovich potential, or simply Kantorovich potential, for the couple μ, ν. A c-convex function φ is called c-convex Kantorovich potential if $-\varphi$ is a c-concave Kantorovich potential.*

Observe that c-concave Kantorovich potentials are related to the transport problem in the following two different (but clearly related) ways:

- as c-concave functions whose superdifferential contains the support of optimal plans, according to Theorem 1.13,

- as maximizing functions, together with their c_+-tranforms, in the dual problem.

1.4 Existence of optimal maps

The problem of existence of optimal transport maps consists in looking for optimal plan $\boldsymbol{\gamma}$ which are induced by a map $T : X \to Y$, i.e. plans $\boldsymbol{\gamma}$ which are equal to $(Id, T)_\# \mu$, for $\mu := \pi^X_\# \boldsymbol{\gamma}$ and some measurable map T. As we discussed in the first section, in general this problem has no answer, as it may very well be the case when, for given $\mu \in \mathscr{P}(X)$, $\nu \in \mathscr{P}(Y)$, there is no transport map at all from μ to ν. Still, since we know that (1.2) holds when μ has no atom, it is possible that under some additional assumptions on the starting measure μ and on the cost function c, optimal transport maps exist.

To formulate the question differently: given μ, ν and the cost function c, is that true that at least one optimal plan $\boldsymbol{\gamma}$ is induced by a map?

Let us start observing that thanks to Theorem 1.13, the answer to this question relies in a natural way on the analysis of the properties of c-monotone sets, to see how far are they from being graphs. Indeed:

Lemma 1.20 *Let $\boldsymbol{\gamma} \in \mathcal{A}dm(\mu, \nu)$. Then $\boldsymbol{\gamma}$ is induced by a map if and only if there exists a $\boldsymbol{\gamma}$-measurable set $\Gamma \subset X \times Y$ where $\boldsymbol{\gamma}$ is concentrated, such that for μ-a.e. x there exists only one $y = T(x) \in Y$ such that $(x, y) \in \Gamma$. In this case $\boldsymbol{\gamma}$ is induced by the map T.*

Proof The *if* part is obvious. For the *only if*, let Γ be as in the statement of the lemma. Possibly removing from Γ a product $N \times Y$, with N μ-negligible, we can assume that Γ is a graph, and denote by T the corresponding map. By the inner regularity of measures, it is easily seen that we can also assume $\Gamma = \cup_n \Gamma_n$ to be σ-compact. Under this assumption the domain of T (i.e. the projection of Γ on X) is σ-compact, hence Borel, and the restriction of T to the compact set $\pi_X(\Gamma_n)$ is continuous. It follows that T is a Borel map. Since $y = T(x)$ $\boldsymbol{\gamma}$-a.e. in $X \times Y$ we conclude that

$$\int \phi(x, y)\, d\boldsymbol{\gamma}(x, y) = \int \phi(x, T(x)) d\boldsymbol{\gamma}(x, y) = \int \phi(x, T(x)) d\mu(x),$$

so that $\boldsymbol{\gamma} = (Id \times T)_\# \mu$. □

Thus the point is the following. We know by Theorem 1.13 that optimal plans are concentrated on c-cyclically monotone sets, still from Theorem 1.13 we know that c-cyclically monotone sets are obtained by taking the c-superdifferential of a c-concave function. Hence from the lemma above what we need to understand is "how often" the c-superdifferential of a c-concave function is single valued.

There is no general answer to this question, but many particular cases can be studied. Here we focus on two special and very important situations:

- $X = Y = \mathbb{R}^d$ and $c(x,y) = |x-y|^2/2$,
- $X = Y = M$, where M is a Riemannian manifold, and $c(x,y) = d^2(x,y)/2$, d being the Riemannian distance.

Let us start with the case $X = Y = \mathbb{R}^d$ and $c(x,y) = |x-y|^2/2$. In this case there is a simple characterization of c-concavity and c-superdifferential:

Proposition 1.21 *Let $\varphi : \mathbb{R}^d \to \mathbb{R} \cup \{-\infty\}$. Then φ is c-concave if and only if $x \mapsto \overline{\varphi}(x) := |x|^2/2 - \varphi(x)$ is convex and lower semicontinuous. In this case $y \in \partial^{c+}\varphi(x)$ if and only if $y \in \partial^-\overline{\varphi}(x)$.*

Proof Observe that

$$
\begin{aligned}
\varphi(x) = \inf_y \frac{|x-y|^2}{2} - \psi(y) &\Leftrightarrow \varphi(x) = \inf_y \frac{|x|^2}{2} + \langle x, -y\rangle + \frac{|y|^2}{2} - \psi(y) \\
&\Leftrightarrow \varphi(x) - \frac{|x|^2}{2} = \inf_y \langle x, -y\rangle + \left(\frac{|y|^2}{2} - \psi(y)\right) \\
&\Leftrightarrow \overline{\varphi}(x) = \sup_y \langle x, y\rangle - \left(\frac{|y|^2}{2} - \psi(y)\right),
\end{aligned}
$$

which proves the first claim. For the second observe that

$$
\begin{aligned}
y \in \partial^{c+}\varphi(x) &\Leftrightarrow \begin{cases} \varphi(x) = |x-y|^2/2 - \varphi^{c+}(y), \\ \varphi(z) \le |z-y|^2/2 - \varphi^{c+}(y), & \forall z \in \mathbb{R}^d \end{cases} \\
&\Leftrightarrow \begin{cases} \varphi(x) - |x|^2/2 = \langle x, -y\rangle + |y|^2/2 - \varphi^{c+}(y), \\ \varphi(z) - |z|^2/2 \le \langle z, -y\rangle + |y|^2/2 - \varphi^{c+}(y), & \forall z \in \mathbb{R}^d \end{cases} \\
&\Leftrightarrow \varphi(z) - |z|^2/2 \le \varphi(x) - |x|^2/2 + \langle z - x, -y\rangle \qquad \forall z \in \mathbb{R}^d \\
&\Leftrightarrow -y \in \partial^+(\varphi - |\cdot|^2/2)(x) \\
&\Leftrightarrow y \in \partial^-\overline{\varphi}(x)
\end{aligned}
$$

□

Therefore in this situation being concentrated on the c-superdifferential of a c-concave map means being concentrated on the (graph of) the subdifferential of a convex function.

Remark 1.22 (Perturbations of the identity via smooth gradients are optimal) An immediate consequence of the above proposition is the fact that if $\psi \in C_c^\infty(\mathbb{R}^d)$, then there exists $\bar{\varepsilon} > 0$ such that $Id + \varepsilon\nabla\psi$ is an optimal map for any $|\varepsilon| \le \bar{\varepsilon}$. Indeed, it is sufficient to take $\bar{\varepsilon}$ such that $-Id \le \bar{\varepsilon}\nabla^2\psi \le Id$. With this choice, the map $x \mapsto |x|^2/2 + \varepsilon\psi(x)$ is convex for any $|\varepsilon| \le \bar{\varepsilon}$, and thus its gradient is an optimal map. ■

Proposition 1.21 reduced the problem of understanding when there exists optimal maps reduces to the problem of convex analysis of understanding how the set of non differentiability points of a convex function is made. This latter problem has a known answer; in order to state it, we need the following definition:

Definition 1.23 ($c-c$ hypersurfaces) *A set $E \subset \mathbb{R}^d$ is called $c-c$ hypersurface*[1] *if, in a suitable system of coordinates, it is the graph of the difference of two real valued convex functions, i.e. if there exists convex functions f, $g : \mathbb{R}^{d-1} \to \mathbb{R}$ such that*

$$
E = \left\{(y,t) \in \mathbb{R}^d \,:\, y \in \mathbb{R}^{d-1},\, t \in \mathbb{R},\, t = f(y) - g(y)\right\}.
$$

[1] here $c-c$ stands for 'convex minus convex' and has nothing to do with the c we used to indicate the cost function

Then it holds the following theorem, which we state without proof:

Theorem 1.24 (Structure of sets of non differentiability of convex functions) *Let $A \subset \mathbb{R}^d$. Then there exists a convex function $\overline{\varphi} : \mathbb{R}^d \to \mathbb{R}$ such that A is contained in the set of points of non differentiability of $\overline{\varphi}$ if and only if A can be covered by countably many $c - c$ hypersurfaces.*

We give the following definition:

Definition 1.25 (Regular measures on $\mathbb{R}^d$) *A measure $\mu \in \mathscr{P}(\mathbb{R}^d)$ is called* regular *provided $\mu(E) = 0$ for any $c - c$ hypersurface $E \subset \mathbb{R}^d$.*

Observe that absolutely continuous measures and measures which give 0 mass to Lipschitz hypersurfaces are automatically regular (because convex functions are locally Lipschitz, thus a $c - c$ hypersurface is a locally Lipschitz hypersurface).

Now we can state the result concerning existence and uniqueness of optimal maps:

Theorem 1.26 (Brenier) *Let $\mu \in \mathscr{P}(\mathbb{R}^d)$ be such that $\int |x|^2 d\mu(x)$ is finite. Then the following are equivalent:*

i) for every $\nu \in \mathscr{P}(\mathbb{R}^d)$ with $\int |x|^2 d\nu(x) < \infty$ there exists only one transport plan from μ to ν and this plan is induced by a map T,

ii) μ is regular.

If either (i) or (ii) hold, the optimal map T can be recovered by taking the gradient of a convex function.

Proof

$(\mathbf{ii}) \Rightarrow (\mathbf{i})$ **and the last statement**. Take $a(x) = b(x) = |x|^2$ in the statement of Theorem 1.13. Then our assumptions on μ, ν guarantees that the bound (1.4) holds. Thus the conclusions of Theorems 1.13 and 1.17 are true as well. Using Remark 1.18 we know that for any c-concave Kantorovich potential φ and any optimal plan $\boldsymbol{\gamma} \in \mathcal{Opt}(\mu, \nu)$ it holds $\mathrm{supp}(\boldsymbol{\gamma}) \subset \partial^{c+}\varphi$. Now from Proposition 1.21 we know that $\overline{\varphi} := |\cdot|^2/2 - \varphi$ is convex and that $\partial^c \varphi = \partial^- \overline{\varphi}$. Here we use our assumption on μ: since $\overline{\varphi}$ is convex, we know that the set E of points of non differentiability of $\overline{\varphi}$ is μ-negligible. Therefore the map $\nabla\overline{\varphi} : \mathbb{R}^d \to \mathbb{R}^d$ is well defined μ-a.e. and every optimal plan must be concentrated on its graph. Hence the optimal plan is unique and induced by the gradient of the convex function $\overline{\varphi}$.

$(\mathbf{ii}) \Rightarrow (\mathbf{i})$. We argue by contradiction and assume that there is some convex function $\overline{\varphi} : \mathbb{R}^d \to \mathbb{R}$ such that the set E of points of non differentiability of $\overline{\varphi}$ has positive μ measure. Possibly modifying $\overline{\varphi}$ outside a compact set, we can assume that it has linear growth at infinity. Now define the two maps:

$$T(x) := \text{the element of smallest norm in } \partial^-\overline{\varphi}(x),$$
$$S(x) := \text{the element of biggest norm in } \partial^-\overline{\varphi}(x),$$

and the plan

$$\boldsymbol{\gamma} := \frac{1}{2}\big((Id, T)_\# \mu + (Id, S)_\# \mu\big).$$

The fact that $\overline{\varphi}$ has linear growth, implies that $\nu := \pi^Y_\# \boldsymbol{\gamma}$ has compact support. Thus in particular $\int |x|^2 d\nu(x) < \infty$. The contradiction comes from the fact that $\boldsymbol{\gamma} \in \mathcal{Adm}(\mu, \nu)$ is c-cyclically monotone (because of Proposition 1.21), and thus optimal. However, it is not induced by a map, because $T \neq S$ on a set of positive μ measure (Lemma 1.20). $\square$

The question of *regularity* of the optimal map is very delicate. In general it is only of bounded variation (BV in short), since monotone maps always have this regularity property, and discontinuities can occur: just think to the case in which the support of the starting measure is connected, while the one of the arrival measure is not. It turns out that connectedness is not sufficient to prevent discontinuities, and that if we want some regularity, we have to impose a convexity restriction on $\operatorname{supp}\nu$. The following result holds:

Theorem 1.27 (Regularity theorem) *Assume $\Omega_1, \Omega_2 \subset \mathbb{R}^d$ are two bounded and connected open sets, $\mu = \rho\mathcal{L}^d|_{\Omega_1}$, $\nu = \eta\mathcal{L}^d|_{\Omega_2}$ with $0 < c \le \rho, \eta \le C$ for some $c, C \in \mathbb{R}$. Assume also that Ω_2 is convex. Then the optimal transport map T belongs to $C^{0,\alpha}(\Omega_1)$ for some $\alpha < 1$. In addition, the following implication holds:*

$$\rho \in C^{0,\alpha}(\Omega_1),\ \ \eta \in C^{0,\alpha}(\Omega_2) \qquad \Longrightarrow \qquad T \in C^{1,\alpha}(\Omega_1).$$

The convexity assumption on Ω_2 is needed to show that the convex function φ whose gradient provides the optimal map T is a *viscosity* solution of the Monge-Ampere equation

$$\rho^1(x) = \rho^2(\nabla\varphi(x)) \det(\nabla^2\varphi(x)),$$

and then the regularity theory for Monge-Ampere, developed by Caffarelli and Urbas, applies.

As an application of Theorem 1.26 we discuss the question of *polar factorization* of vector fields on $\mathbb{R}^d$. Let $\Omega \subset \mathbb{R}^d$ be a bounded domain, denote by μ_Ω the normalized Lebesgue measure on Ω and consider the space

$$S(\Omega) := \{\text{Borel map } s : \Omega \to \Omega \ : \ s_{\#}\mu_\Omega = \mu_\Omega\}.$$

The following result provides a (nonlinear) projection on the (nonconvex) space $S(\Omega)$.

Proposition 1.28 (Polar factorization) *Let $S \in L^2(\mu_\Omega; \mathbb{R}^n)$ be such that $\nu := S_{\#}\mu$ is regular (Definition 1.25). Then there exist unique $s \in S(\Omega)$ and $\nabla\varphi$, with φ convex, such that $S = (\nabla\varphi) \circ s$. Also, s is the unique minimizer of*

$$\int |S - \tilde{s}|^2 d\mu,$$

among all $\tilde{s} \in S(\Omega)$.

Proof By assumption, we know that both μ_Ω and ν are regular measures with finite second moment. We claim that

$$\inf_{\tilde{s}\in S(\Omega)} \int |S - \tilde{s}|^2 d\mu = \min_{\boldsymbol{\gamma}\in \mathcal{Adm}(\mu,\nu)} \int |x - y|^2 d\boldsymbol{\gamma}(x, y). \tag{1.7}$$

To see why, associate to each $\tilde{s} \in S(\Omega)$ the plan $\boldsymbol{\gamma}_{\tilde{s}} := (\tilde{s}, S)_{\#}\mu$ which clearly belongs to $\mathcal{Adm}(\mu_\Omega, \nu)$. This gives inequality $\ge$. Now let $\overline{\boldsymbol{\gamma}}$ be the unique optimal plan and apply Theorem 1.26 twice to get that

$$\overline{\boldsymbol{\gamma}} = (Id, \nabla\varphi)_{\#}\mu_\Omega = (\nabla\tilde{\varphi}, Id)_{\#}\nu,$$

for appropriate convex functions $\varphi, \tilde{\varphi}$, which therefore satisfy $\nabla\varphi \circ \nabla\tilde{\varphi} = Id$ μ-a.e.. Define $s := \nabla\tilde{\varphi} \circ S$. Then $s_{\#}\mu_\Omega = \mu_\Omega$ and thus $s \in S(\Omega)$. Also, $S = \nabla\varphi \circ s$ which proves the existence of the polar factorization. The identity

$$\int |x - y|^2 d\boldsymbol{\gamma}_s(x, y) = \int |s - S|^2 d\mu_\Omega = \int |\nabla\tilde{\varphi} \circ S - S|^2 d\mu_\Omega = \int |\nabla\tilde{\varphi} - Id|^2 d\nu$$
$$= \min_{\boldsymbol{\gamma}\in \mathcal{Adm}(\mu,\nu)} \int |x - y|^2 d\boldsymbol{\gamma}(x, y),$$

shows inequality $\leq$ in (1.7) and the uniqueness of the optimal plan ensures that s is the unique minimizer.

To conclude we need to show uniqueness of the polar factorization. Assume that $S = (\nabla\overline{\varphi}) \circ \overline{s}$ is another factorization and notice that $\nabla\overline{\varphi}_{\#}\mu_\Omega = (\nabla\overline{\varphi} \circ \overline{s})_{\#}\mu_\Omega = \nu$. Thus the map $\nabla\overline{\varphi}$ is a transport map from μ_Ω to ν and is the gradient of a convex function. By Proposition 1.21 and Theorem 1.13 we deduce that $\nabla\overline{\varphi}$ is the optimal map. Hence $\nabla\overline{\varphi} = \nabla\varphi$ and the proof is achieved. $\square$

Remark 1.29 (Polar factorization vs Helmholtz decomposition) The classical Helmoltz decomposition of vector fields can be seen as a linearized version of the polar factorization result, which therefore can be though as a generalization of the former.

To see why, assume that Ω and all the objects considered are smooth (the arguments hereafter are just formal). Let $u : \Omega \to \mathbb{R}^d$ be a vector field and apply the polar factorization to the map $S_\varepsilon := Id + \varepsilon u$ with $|\varepsilon|$ small. Then we have $S_\varepsilon = (\nabla\varphi_\varepsilon) \circ s_\varepsilon$ and both $\nabla\varphi_\varepsilon$ and s_ε will be perturbation of the identity, so that

$$\begin{aligned}\nabla\varphi_\varepsilon &= Id + \varepsilon v + o(\varepsilon),\\ s_\varepsilon &= Id + \varepsilon w + o(\varepsilon).\end{aligned}$$

The question now is: which information is carried on v, w from the properties of the polar factorization? At the level of v, from the fact that $\nabla \times (\nabla\varphi_\varepsilon) = 0$ we deduce $\nabla \times v = 0$, which means that v is the gradient of some function p. On the other hand, the fact that s_ε is measure preserving implies that w satisfies $\nabla \cdot (w\chi_\Omega) = 0$ in the sense of distributions: indeed for any smooth $f : \mathbb{R}^d \to \mathbb{R}$ it holds

$$0 = \frac{d}{d\varepsilon}_{|\varepsilon=0} \int f \, d(s_\varepsilon)_{\#}\mu_\Omega = \frac{d}{d\varepsilon}_{|\varepsilon=0} \int f \circ s_\varepsilon \, d\mu_\Omega = \int \nabla f \cdot w \, d\mu_\Omega.$$

Then from the identity $(\nabla\varphi_\varepsilon) \circ s_\varepsilon = Id + \varepsilon(\nabla p + w) + o(\varepsilon)$ we can conclude that

$$u = \nabla p + w.$$

$\blacksquare$

We now turn to the case $X = Y = M$, with M smooth Riemannian manifold, and $c(x, y) = d^2(x, y)/2$, d being the Riemannian distance on M. For simplicity, we will assume that M is compact and with no boundary, but everything holds in more general situations.

The underlying ideas of the foregoing discussion are very similar to the ones of the case $X = Y = \mathbb{R}^d$, the main difference being that there is no more the correspondence given by Proposition 1.21 between c-concave functions and convex functions, as in the Euclidean case. Recall however that the concepts of semiconvexity (i.e. second derivatives bounded from below) and semiconcavity make sense also on manifolds, since these properties can be read locally and changes of coordinates are smooth.

In the next proposition we will use the fact that on a compact and smooth Riemannian manifold, the functions $x \mapsto d^2(x, y)$ are uniformly Lipschitz and uniformly semiconcave in $y \in M$ (i.e. the second derivative along a unit speed geodesic is bounded above by a universal constant depending only on M, see e.g. the third appendix of Chapter 10 of [80] for the simple proof).

Proposition 1.30 *Let M be a smooth, compact Riemannian manifold without boundary. Let $\varphi : M \to \mathbb{R} \cup \{-\infty\}$ be a c-concave function not identically equal to $-\infty$. Then φ is Lipschitz, semiconcave and real valued. Also, assume that $y \in \partial^{c+}\varphi(x)$. Then $\exp_x^{-1}(y) \subset -\partial^+\varphi(x)$. Conversely, if φ is differentiable at x, then $\exp_x(-\nabla\varphi(x)) \in \partial^{c+}\varphi(x)$.*

Proof The fact that φ is real valued follows from the fact that the cost function $d^2(x,y)/2$ is uniformly bounded in $x, y \in M$. Smoothness and compactness ensure that the functions $d^2(\cdot,y)/2$ are uniformly Lipschitz and uniformly semiconcave in $y \in M$, this gives that φ is Lipschitz and semiconcave.

Now pick $y \in \partial^{c+}\varphi(x)$ and $\mathrm{v} \in \exp_x^{-1}(y)$. Recall that $-\mathrm{v}$ belongs to the superdifferential of $d^2(\cdot,y)/2$ at x, i.e.

$$\frac{d^2(z,y)}{2} \le \frac{d^2(x,y)}{2} - \left\langle \mathrm{v}, \exp_x^{-1}(z)\right\rangle + o(d(x,z)).$$

Thus from $y \in \partial^{c+}\varphi(x)$ we have

$$\varphi(z) - \varphi(x) \overset{(1.3)}{\le} \frac{d^2(z,y)}{2} - \frac{d^2(x,y)}{2} \le \left\langle -\mathrm{v}, \exp_x^{-1}(z)\right\rangle + o(d(x,z)),$$

that is $-\mathrm{v} \in \partial^+\varphi(x)$.

To prove the converse implication, it is enough to show that the c-superdifferential of φ at x is non empty. To prove this, use the c-concavity of φ to find a sequence $(y_n) \subset M$ such that

$$\begin{aligned}\varphi(x) &= \lim_{n\to\infty} \frac{d^2(x,y_n)}{2} - \varphi^{c+}(y_n),\\ \varphi(z) &\le \frac{d^2(z,y_n)}{2} - \varphi^{c+}(y_n), \qquad \forall z \in M,\ n \in \mathbb{N}.\end{aligned}$$

By compactness we can extract a subsequence converging to some $y \in M$. Then from the continuity of $d^2(z,\cdot)/2$ and $\varphi^{c+}(\cdot)$ it is immediate to verify that $y \in \partial^{c+}\varphi(x)$. □

Remark 1.31 The converse implication in the previous proposition is *false* if one doesn't assume φ to be differentiable at x: i.e., it is *not* true in general that $\exp_x(-\partial^+\varphi(x)) \subset \partial^{c+}\varphi(x)$. ■

From this proposition, and following the same ideas used in the Euclidean case, we give the following definition:

Definition 1.32 (Regular measures in $\mathscr{P}(M)$) *We say that $\mu \in \mathscr{P}(M)$ is regular provided it vanishes on the set of points of non differentiability of ψ for any semiconvex function $\psi : M \to \mathbb{R}$.*

The set of points of non differentiability of a semiconvex function on M can be described as in the Euclidean case by using local coordinates. For most applications it is sufficient to keep in mind that absolutely continuous measures (w.r.t. the volume measure) and even measures vanishing on Lipschitz hypersurfaces are regular.

By Proposition 1.30, we can derive a result about existence and characterization of optimal transport maps in manifolds which closely resembles Theorem 1.26:

Theorem 1.33 (McCann) *Let M be a smooth, compact Riemannian manifold without boundary and $\mu \in \mathscr{P}(M)$. Then the following are equivalent:*

i) for every $\nu \in \mathscr{P}(M)$ there exists only one transport plan from μ to ν and this plan is induced by a map T,

ii) μ is regular.

If either (i) or (ii) hold, the optimal map T can be written as $x \mapsto \exp_x(-\nabla\varphi(x))$ for some c-concave function $\varphi : M \to \mathbb{R}$.

Proof
(ii) $\Rightarrow$ **(i) and the last statement**. Pick $\nu \in \mathscr{P}(M)$ and observe that, since $d^2(\cdot,\cdot)/2$ is uniformly bounded, condition (1.4) surely holds. Thus from Theorem 1.13 and Remark 1.15 we get that any optimal plan $\boldsymbol{\gamma} \in Opt(\mu,\nu)$ must be concentrated on the c-superdifferential of a c-concave function φ. By Proposition 1.30 we know that φ is semiconcave, and thus differentiable μ-a.e. by our assumption on μ. Therefore $x \mapsto T(x) := \exp_x(-\nabla\varphi(x))$ is well defined μ-a.e. and its graph must be of full $\boldsymbol{\gamma}$-measure for any $\boldsymbol{\gamma} \in Opt(\mu,\nu)$. This means that $\boldsymbol{\gamma}$ is unique and induced by T.
(i) $\Rightarrow$ **(ii)**. Argue by contradiction and assume that there exists a semiconcave function f whose set of points of non differentiability has positive μ measure. Use Lemma 1.34 below to find $\varepsilon > 0$ such that $\varphi := \varepsilon f$ is c-concave and satisfies: $\mathrm{v} \in \partial^+\varphi(x)$ if and only $\exp_x(-\mathrm{v}) \in \partial^{c+}\varphi(x)$. Then conclude the proof as in Theorem 1.26. $\square$

Lemma 1.34 *Let M be a smooth, compact Riemannian manifold without boundary and $\varphi : M \to \mathbb{R}$ semiconcave. Then for $\varepsilon > 0$ sufficiently small the function $\varepsilon\varphi$ is c-concave and it holds $\mathrm{v} \in \partial^+(\varepsilon\varphi)(x)$ if and only $\exp_x(-\mathrm{v}) \in \partial^{c+}(\varepsilon\varphi)(x)$.*

Proof We start with the following claim: there exists $\varepsilon > 0$ such that for every $x_0 \in M$ and every $\mathrm{v} \in \partial^+\varphi(x_0)$ the function
$$x \mapsto \varepsilon\varphi(x) - \frac{d^2(x, \exp_{x_0}(-\varepsilon\mathrm{v}))}{2}$$
has a global maximum at $x = x_0$.

Use the smoothness and compactness of M to find $r > 0$ such that $d^2(\cdot,\cdot)/2 : \{(x,y) : d(x,y) < r\} \to \mathbb{R}$ is C^∞ and satisfies $\nabla^2 d^2(\cdot,y)/2 \geq cId$, for every $y \in M$, with $c > 0$ independent on y. Now observe that since φ is semiconcave and real valued, it is Lipschitz. Thus, for $\varepsilon_0 > 0$ sufficiently small it holds $\varepsilon_0|\mathrm{v}| < r/3$ for any $\mathrm{v} \in \partial^+\varphi(x)$ and any $x \in M$. Also, since φ is bounded, possibly decreasing the value of ε_0 we can assume that
$$\varepsilon_0|\varphi(x)| \leq \frac{r^2}{12}.$$
Fix $x_0 \in M$, $\mathrm{v} \in \partial^+\varphi(x_0)$ and let $y_0 := \exp_{x_0}(-\varepsilon_0\mathrm{v})$. We claim that for ε_0 chosen as above, the maximum of $\varepsilon_0\varphi - d^2(\cdot,y_0)/2$, cannot lie outside $B_r(x_0)$. Indeed, if $d(x,x_0) \geq r$ we have $d(x,y_0) > 2r/3$ and thus:
$$\varepsilon_0\varphi(x) - \frac{d^2(x,y_0)}{2} < \frac{r^2}{12} - \frac{2r^2}{9} = -\frac{r^2}{12} - \frac{r^2}{18} \leq \varepsilon_0\varphi(x_0) - \frac{d^2(x_0,y_0)}{2}.$$
Thus the maximum must lie in $B_r(x_0)$. Recall that in this ball, the function $d^2(\cdot,y_0)$ is C^∞ and satisfies $\nabla^2(d^2(\cdot,y_0)/2) \geq cId$, thus it holds
$$\nabla^2\left(\varepsilon_0\varphi(\cdot) - \frac{d^2(\cdot,y_0)}{2}\right) \leq (\varepsilon_0\lambda - c)Id,$$
where $\lambda \in \mathbb{R}$ is such that $\nabla^2\varphi \leq \lambda Id$ on the whole of M. Thus decreasing if necessary the value of ε_0 we can assume that
$$\nabla^2\left(\varepsilon_0\varphi(\cdot) - \frac{d^2(\cdot,y_0)}{2}\right) < 0 \qquad \text{on } B_r(x_0),$$
which implies that $\varepsilon_0\varphi(\cdot) - d^2(\cdot,y_0)/2$ admits a unique point $x \in B_r(x_0)$ such that $0 \in \partial^+(\varphi - d^2(\cdot,y_0)/2)(x)$, which therefore is the unique maximum. Since $\nabla\frac{1}{2}d^2(\cdot,y_0)(x_0) = \varepsilon_0\mathrm{v} \in \partial^+(\varepsilon_0\varphi)(x_0)$, we conclude that x_0 is the unique global maximum, as claimed.

Now define the function $\psi : M \to \mathbb{R} \cup \{-\infty\}$ by

$$\psi(y) := \inf_{x \in M} \frac{d^2(x, y)}{2} - \varepsilon_0 \varphi(x),$$

if $y = \exp_x(-\varepsilon_0 \mathrm{v})$ for some $x \in M$, $\mathrm{v} \in \partial^+ \varphi(x)$, and $\psi(y) := -\infty$ otherwise. By definition we have

$$\varepsilon_0 \varphi(x) \leq \frac{d^2(x, y)}{2} - \psi(y), \qquad \forall x, y \in M,$$

and the claim proved ensures that if $y_0 = \exp_{x_0}(-\varepsilon_0 \mathrm{v}_0)$ for $x_0 \in M$, $\mathrm{v}_0 \in \partial^+ \varphi(x_0)$ the $\inf$ in the definition of $\psi(y_0)$ is realized at $x = x_0$ and thus

$$\varepsilon_0 \varphi(x_0) = \frac{d^2(x_0, y_0)}{2} - \psi(y_0).$$

Hence $\varepsilon_0 \varphi = \psi^{c+}$ and therefore is c-concave. Along the same lines one can easily see that for $y \in \exp_x(-\varepsilon_0 \partial^+ \varphi(x))$ it holds

$$\varepsilon_0 \varphi(x) = \frac{d^2(x, y)}{2} - (\varepsilon_0 \varphi)^{c+}(y),$$

i.e. $y \in \partial^{c+}(\varepsilon_0 \varphi)(x_0)$. Thus we have $\partial^{c+}(\varepsilon_0 \varphi) \supset \exp(-\partial^+(\varepsilon \varphi))$. Since the other inclusion has been proved in Proposition 1.30 the proof is finished. □

Remark 1.35 With the same notation of Theorem 1.33, recall that we know that the c-concave function φ whose c-superdifferential contains the graph of any optimal plan from μ to ν is differentiable μ-a.e. (for regular μ). Fix x_0 such that $\nabla \varphi(x_0)$ exists, let $y_0 := \exp_{x_0}(-\nabla \varphi(x_0)) \in \partial^{c+} \varphi(x_0)$ and observe that from

$$\frac{d^2(x, y_0)}{2} - \frac{d^2(x_0, y_0)}{2} \geq \varphi(x) - \varphi(x_0),$$

we deduce that $\nabla \varphi(x_0)$ belongs to the *sub*differential of $d^2(\cdot, y_0)/2$ at x_0. Since we know that $d^2(\cdot, y_0)/2$ always have non empty superdifferential, we deduce that it must be differentiable at x_0. In particular, *there exists only one geodesic connecting x_0 to y_0*. Therefore if μ is regular, not only there exists a unique optimal transport map T, but also for μ-a.e. x there is only one geodesic connecting x to $T(x)$. ■

The question of regularity of optimal maps on manifolds is much more delicate than the corresponding question on $\mathbb{R}^d$, even if one wants to get only the continuity. We won't enter into the details of the theory, we just give an example showing the difficulty that can arise in a curved setting. The example will show a smooth compact manifold, and two measures absolutely continuous with positive and smooth densities, such that the optimal transport map is discontinuous. We remark that similar behaviors occur as soon as M has one point and one sectional curvature at that point which is strictly negative. Also, even if one assumes that the manifold has non negative sectional curvature everywhere, this is not enough to guarantee continuity of the optimal map: what comes into play in this setting is the Ma-Trudinger-Wang tensor, an object which we will not study.

Example 1.36 Let $M \subset \mathbb{R}^3$ be a smooth surface which has the following properties:

- M is symmetric w.r.t. the x axis and the y axis,
- M crosses the line $(x, y) = (0, 0)$ at two points, namely O, O'.

- the curvature of M at O is negative.

These assumptions ensure that we can find $a, b > 0$ such that for some z_a, z_b the points

$$\begin{aligned} A &:= (a, 0, z_a), \\ A' &:= (-a, 0, z_a), \\ B &:= (0, b, z_b), \\ B' &:= (0, -b, z_b), \end{aligned}$$

belong to M and

$$d^2(A, B) > d^2(A, O) + d^2(O, B),$$

d being the intrinsic distance on M. By continuity and symmetry, we can find $\varepsilon > 0$ such that

$$d^2(x, y) > d^2(x, O) + d^2(O, y), \qquad \forall x \in B_\varepsilon(A) \cup B_\varepsilon(A'),\ y \in B_\varepsilon(B) \cup B_\varepsilon(B'). \tag{1.8}$$

Now let f (resp. g) be a smooth probability density everywhere positive and symmetric w.r.t. the x, y axes such that $\int_{B_\varepsilon(A)\cup B_\varepsilon(A')} f \, d\mathrm{vol} > \frac{1}{2}$ (resp. $\int_{B_\varepsilon(B)\cup B_\varepsilon(B')} g \, d\mathrm{vol} > \frac{1}{2}$), and let T (resp. T') be the optimal transport map from $f\mathrm{vol}$ to $g\mathrm{vol}$ (resp. from $g\mathrm{vol}$ to $f\mathrm{vol}$).

We claim that either T or T' is discontinuous and argue by contradiction. Suppose that both are continuous and observe that by the symmetry of the optimal transport problem it must hold $T'(x) = T^{-1}(x)$ for any $x \in M$. Again by the symmetry of M, f, g, the point $T(O)$ must be invariant under the symmetries around the x and y axes. Thus it is either $T(O) = O$ or $T(O) = O'$, and similarly, $T'(O') \in \{O, O'\}$.

We claim that it must hold $T(O) = O$. Indeed otherwise either $T(O) = O'$ and $T(O') = O$, or $T(O) = O'$ and $T(O') = O'$. In the first case the two couples (O, O') and (O', O) belong to the support of the optimal plan, and thus by cyclical monotonicity it holds

$$d^2(O, O') + d^2(O', O) \le d^2(O, O) + d^2(O', O') = 0,$$

which is absurdum.

In the second case we have $T'(x) \neq O$ for all $x \in M$, which, by continuity and compactness implies $d(T'(M), O) > 0$. This contradicts the fact that f is positive everywhere and $T'_\#(g\mathrm{vol}) = f\mathrm{vol}$.

Thus it holds $T(O) = O$. Now observe that by construction there must be some mass transfer from $B_\varepsilon(A) \cup B_\varepsilon(A')$ to $B_\varepsilon(B) \cup B_\varepsilon(B')$, i.e. we can find $x \in B_\varepsilon(A) \cup B_\varepsilon(A')$ and $y \in B_\varepsilon(B) \cup B_\varepsilon(B')$ such that (x, y) is in the support of the optimal plan. Since (O, O) is the support of the optimal plan as well, by cyclical monotonicity it must hold

$$d^2(x, y) + d^2(O, O) \le d^2(x, O) + d^2(O, y),$$

which contradicts (1.8). ■

1.5 Bibliographical notes

G. Monge's original formulation of the transport problem ([66]) was concerned with the case $X = Y = \mathbb{R}^d$ and $c(x, y) = |x - y|$, and L. V. Kantorovich's formulation appeared first in [49].

The equality (1.2), saying that the infimum of the Monge problem equals the minimum of Kantorovich one, has been proved by W. Gangbo (Appendix A of [41]) and the first author (Theorem 2.1 in [4]) in particular cases, and then generalized by A. Pratelli [68].

In [50] L. V. Kantorovich introduced the dual problem, and later L. V. Kantorovich and G. S. Rubinstein [51] further investigated this duality for the case $c(x, y) = d(x, y)$. The fact that the study of the dual problem can lead to important informations for the transport problem has been investigated by several authors, among others M. Knott and C. S. Smith [52] and S. T. Rachev and L. Rüschendorf [69], [71].

The notions of cyclical monotonicity and its relation with subdifferential of convex function have been developed by Rockafellar in [70]. The generalization to c-cyclical monotonicity and to c-sub/super differential of c-convex/concave functions has been studied, among others, by Rüschendorf [71].

The characterization of the set of non differentiability of convex functions is due to Zajíček ([83], see also the paper by G. Alberti [2] and the one by G. Alberti and the first author [3])

Theorem 1.26 on existence of optimal maps in $\mathbb{R}^d$ for the cost=distance-squared is the celebrated result of Y. Brenier, who also observed that it implies the polar factorization result 1.28 ([18], [19]). Brenier's ideas have been generalized in many directions. One of the most notable one is R. McCann's theorem 1.33 concerning optimal maps in Riemannian manifolds for the case cost=squared distance ([64]). R. McCann also noticed that the original hypothesis in Brenier's theorem, which was $\mu \ll \mathcal{L}^d$, can be relaxed into 'μ gives 0 mass to Lipschitz hypersurfaces'. In [42] W. Gangbo and R. McCann pointed out that to get existence of optimal maps in $\mathbb{R}^d$ with $c(x, y) = |x - y|^2/2$ it is sufficient to ask to the measure μ to be regular in the sense of the Definition 1.25. The sharp version of Brenier's and McCann's theorems presented here, where the necessity of the regularity of μ is also proved, comes from a paper of the second author of these notes ([46]).

Other extensions of Brenier's result are:

- Infinite-dimensional Hilbert spaces (the authors and Savaré - [6])
- cost functions induced by Lagrangians, Bernard-Buffoni [13], namely

$$c(x, y) := \inf \left\{ \int_0^1 L(t, \gamma(t), \dot{\gamma}(t))\, dt : \ \gamma(0) = x, \ \gamma(1) = y \right\};$$

- Carnot groups and sub-Riemannian manifolds, $c = d^2_{CC}/2$: the first author and S. Rigot ([10]), A. Figalli and L. Rifford ([39]);
- cost functions induced by sub-Riemannian Lagrangians A. Agrachev and P. Lee ([1]).
- Wiener spaces (E, H, γ), D. Feyel- A. S. Üstünel ([36]).
  Here E is a Banach space, $\gamma \in \mathscr{P}(E)$ is Gaussian and H is its Cameron- Martin space, namely

$$H := \{h \in E : \ (\tau_h)_\sharp \gamma \ll \gamma\}.$$

  In this case

$$c(x, y) := \begin{cases} \dfrac{|x - y|_H^2}{2} & \text{if } x - y \in H; \\ +\infty & \text{otherwise.} \end{cases}$$

The issue of regularity of optimal maps would nowadays require a lecture note in its own. A rough statement that one should have in mind is that it is rare to have regular (even just continuous) optimal transport maps. The key Theorem 1.27 is due to L. Caffarelli ([22], [21], [23]).

Example 1.36 is due to G. Loeper ([55]). For the general case of cost=squared distance on a compact Riemannian manifold, it turns out that continuity of optimal maps between two measures with smooth and strictly positive density is strictly related to the positivity of the so-called Ma-Trudinger-Wang tensor ([59]), an object defined taking fourth order derivatives of the distance function. The understanding of the structure of this tensor has been a very active research area in the last years, with contributions coming from X.-N. Ma, N. Trudinger, X.-J. Wang, C. Villani, P. Delanoe, R. McCann, A. Figalli, L. Rifford, H.-Y. Kim and others.

A topic which we didn't discuss at all is the original formulation of the transport problem of Monge: the case $c(x, y) := |x - y|$ on $\mathbb{R}^d$. The situation in this case is much more complicated than the one with $c(x, y) = |x - y|^2/2$ as it is typically not true that optimal plans are unique, or that optimal plans are induced by maps. For example consider on $\mathbb{R}$ any two probability measures μ, ν such that μ is concentrated on the negative numbers and ν on the positive ones. Then one can see that any admissible plan between them is optimal for the cost $c(x, y) = |x - y|$.

Still, even in this case there is existence of optimal maps, but in order to find them one has to use a sort of selection principle. A successful strategy - which has later been applied to a number of different situation - has been proposed by V. N. Sudakov in [77], who used a disintegration principle to reduce the d-dimensional problem to a problem on $\mathbb{R}$. The original argument by V. N. Sudakov was flawed and has been fixed by the first author in [4] in the case of the Euclidean distance. Meanwhile, different proofs of existence of optimal maps have been proposed by L. C.Evans- W. Gangbo ([34]), Trudinger and Wang [78], and L. Caffarelli, M. Feldman and R. McCann [24].

Later, existence of optimal maps for the case $c(x, y) := \|x - y\|$, $\|\cdot\|$ being any norm has been established, at increasing levels of generality, in [9], [28], [27] (containing the most general result, for any norm) and [25].

2 The Wasserstein distance W_2

The aim of this chapter is to describe the properties of the Wasserstein distance W_2 on the space of Borel Probability measures on a given metric space (X, d). This amounts to study the transport problem with cost function $c(x, y) = d^2(x, y)$.

An important characteristic of the Wasserstein distance is that it inherits many interesting geometric properties of the base space (X, d). For this reason we split the foregoing discussion into three sections on which we deal with the cases in which X is: a general Polish space, a geodesic space and a Riemannian manifold.

A word on the notation: when considering product spaces like X^n, with $\pi^i : X^n \to X$ we intend the natural projection onto the i-th coordinate, $i = 1, \dots, n$. Thus, for instance, for $\mu, \nu \in \mathscr{P}(X)$ and $\boldsymbol{\gamma} \in \mathcal{Adm}(\mu, \nu)$ we have $\pi^1_\# \boldsymbol{\gamma} = \mu$ and $\pi^2_\# \boldsymbol{\gamma} = \nu$. Similarly, with $\pi^{i,j} : X^n \to X^2$ we intend the projection onto the i-th and j-th coordinates. And similarly for multiple projections.

2.1 X Polish space

Let (X, d) be a complete and separable metric space.

The distance W_2 is defined as

$$
\begin{aligned}
W_2(\mu, \nu) :=& \sqrt{\inf_{\boldsymbol{\gamma} \in \mathcal{Adm}(\mu,\nu)} \int d^2(x, y) d\boldsymbol{\gamma}(x, y)} \\
=& \sqrt{\int d^2(x, y) d\boldsymbol{\gamma}(x, y)}, \qquad \forall \boldsymbol{\gamma} \in \mathcal{Opt}(\mu, \nu).
\end{aligned}
$$

The natural space to endow with the Wasserstein distance W_2 is the space $\mathscr{P}_2(X)$ of Borel

Probability measures with finite second moment:

$$\mathscr{P}_2(X) := \Big\{ \mu \in \mathscr{P}(X) \ : \ \int d^2(x, x_0) d\mu(x) < \infty \text{ for some, and thus any, } x_0 \in X \Big\}.$$

Notice that if either μ or ν is a Dirac delta, say $\nu = \delta_{x_0}$, then there exists only one plan $\boldsymbol{\gamma}$ in $\mathcal{Adm}(\mu, \nu)$: the plan $\mu \times \delta_{x_0}$, which therefore is optimal. In particular it holds

$$\int d^2(x, x_0) d\mu(x) = W_2^2(\mu, \delta_{x_0}),$$

that is: the second moment is nothing but the squared Wasserstein distance from the corresponding Dirac mass.

We start proving that W_2 is actually a distance on $\mathscr{P}_2(X)$. In order to prove the triangle inequality, we will use the following lemma, which has its own interest:

Lemma 2.1 (Gluing) *Let X, Y, Z be three Polish spaces and let $\boldsymbol{\gamma}^1 \in \mathscr{P}(X \times Y)$, $\boldsymbol{\gamma}^2 \in \mathscr{P}(Y \times Z)$ be such that $\pi^Y_\# \boldsymbol{\gamma}^1 = \pi^Y_\# \boldsymbol{\gamma}^2$. Then there exists a measure $\boldsymbol{\gamma} \in \mathscr{P}(X \times Y \times Z)$ such that*

$$\begin{aligned} \pi^{X,Y}_\# \boldsymbol{\gamma} &= \boldsymbol{\gamma}^1, \\ \pi^{Y,Z}_\# \boldsymbol{\gamma} &= \boldsymbol{\gamma}^2. \end{aligned}$$

Proof Let $\mu := \pi^Y_\# \boldsymbol{\gamma}^1 = \pi^Y_\# \boldsymbol{\gamma}^2$ and use the disintegration theorem to write $d\boldsymbol{\gamma}^1(x, y) = d\mu(y) d\boldsymbol{\gamma}^1_y(x)$ and $d\boldsymbol{\gamma}^2(y, z) = d\mu(y) d\boldsymbol{\gamma}^2_y(z)$. Conclude defining $\boldsymbol{\gamma}$ by

$$d\boldsymbol{\gamma}(x, y, z) := d\mu(y) d(\boldsymbol{\gamma}^1_y \times \boldsymbol{\gamma}^2_y)(x, z).$$

$\square$

Theorem 2.2 (W_2 is a distance) *W_2 is a distance on $\mathscr{P}_2(X)$.*

Proof It is obvious that $W_2(\mu, \mu) = 0$ and that $W_2(\mu, \nu) = W_2(\nu, \mu)$. To prove that $W_2(\mu, \nu) = 0$ implies $\mu = \nu$ just pick an optimal plan $\boldsymbol{\gamma} \in \mathcal{Opt}(\mu, \nu)$ and observe that $\int d^2(x, y) d\boldsymbol{\gamma}(x, y) = 0$ implies that $\boldsymbol{\gamma}$ is concentrated on the diagonal of $X \times X$, which means that the two maps π^1 and π^2 coincide $\boldsymbol{\gamma}$-a.e., and therefore $\pi^1_\# \boldsymbol{\gamma} = \pi^2_\# \boldsymbol{\gamma}$.

For the triangle inequality, we use the gluing lemma to "compose" two optimal plans. Let $\mu_1, \mu_2, \mu_3 \in \mathscr{P}_2(X)$ and let $\boldsymbol{\gamma}^2_1 \in \mathcal{Opt}(\mu_1, \mu_2)$, $\boldsymbol{\gamma}^3_2 \in \mathcal{Opt}(\mu_2, \mu_3)$. By the gluing lemma we know that there exists $\boldsymbol{\gamma} \in \mathscr{P}_2(X^3)$ such that

$$\begin{aligned} \pi^{1,2}_\# \boldsymbol{\gamma} &= \boldsymbol{\gamma}^2_1, \\ \pi^{2,3}_\# \boldsymbol{\gamma} &= \boldsymbol{\gamma}^3_2. \end{aligned}$$

Since $\pi^1_\# \boldsymbol{\gamma} = \mu_1$ and $\pi^3_\# \boldsymbol{\gamma} = \mu_3$, we have $\pi^{1,3}_\# \boldsymbol{\gamma} \in \mathcal{Adm}(\mu_1, \mu_3)$ and therefore from the triangle inequality in $L^2(\boldsymbol{\gamma})$ it holds

$$
\begin{aligned}
W_2(\mu_1,\mu_3) &\leq \sqrt{\int d^2(x_1,x_3)d\pi^{1,3}_{\#}\boldsymbol{\gamma}(x_1,x_3)} = \sqrt{\int d^2(x_1,x_3)d\boldsymbol{\gamma}(x_1,x_2,x_3)} \\
&\leq \sqrt{\int \big(d(x_1,x_2)+d(x_2,x_3)\big)^2 d\boldsymbol{\gamma}(x_1,x_2,x_3)} \\
&\leq \sqrt{\int d^2(x_1,x_2)d\boldsymbol{\gamma}(x_1,x_2,x_3)} + \sqrt{\int d^2(x_2,x_3)d\boldsymbol{\gamma}(x_1,x_2,x_3)} \\
&= \sqrt{\int d^2(x_1,x_2)d\boldsymbol{\gamma}^2_1(x_1,x_2)} + \sqrt{\int d^2(x_2,x_3)d\boldsymbol{\gamma}^3_2(x_2,x_3)} = W_2(\mu_1,\mu_2)+W_2(\mu_2,\mu_3).
\end{aligned}
$$

Finally, we need to prove that W_2 is real valued. Here we use the fact that we restricted the analysis to the space $\mathscr{P}_2(X)$: from the triangle inequality we have

$$
W_2(\mu,\nu) \leq W_2(\mu,\delta_{x_0}) + W_2(\nu,\delta_{x_0}) = \sqrt{\int d^2(x,x_0)d\mu(x)} + \sqrt{\int d^2(x,x_0)d\nu(x)} < \infty.
$$

□

A trivial, yet very useful inequality is:

$$
W_2^2(f_\#\mu, g_\#\mu) \leq \int d_Y^2(f(x),g(x))d\mu(x), \tag{2.1}
$$

valid for any couple of metric spaces X, Y, any $\mu \in \mathscr{P}(X)$ and any couple of Borel maps $f,\, g : X \to Y$. This inequality follows from the fact that $(f,g)_\#\mu$ is an admissible plan for the measures $f_\#\mu$, $g_\#\mu$, and its cost is given by the right hand side of (2.1).

Observe that there is a natural isometric immersion of (X,d) into $(\mathscr{P}_2(X), W_2)$, namely the map $x \mapsto \delta_x$.

Now we want to study the topological properties of $(\mathscr{P}_2(X), W_2)$. To this aim, we introduce the notion of *2-uniform integrability*: $\mathcal{K} \subset \mathscr{P}_2(X)$ is 2-uniformly integrable provided for any $\varepsilon > 0$ and $x_0 \in X$ there exists $R_\varepsilon > 0$ such that

$$
\sup_{\mu\in\mathcal{K}} \int_{X\setminus B_{R_\varepsilon}(x_0)} d^2(x,x_0)d\mu \leq \varepsilon.
$$

Remark 2.3 Let $(X,d_X), (Y,d_Y)$ be Polish and endow $X\times Y$ with the product distance $d^2\big((x_1,y_1),(x_2,y_2)\big) := d_X^2(x_1,x_2) + d_Y^2(y_1,y_2)$. Then the inequality

$$
\begin{aligned}
\int\limits_{(B_R(x_0)\times B_R(y_0))^c} d_X^2(x,x_0)d\boldsymbol{\gamma}(x,y) &= \int\limits_{(B_R(x_0))^c\times Y} d_X^2(x,x_0)d\boldsymbol{\gamma}(x,y) + \int\limits_{B_R(x_0)\times(B_R(y_0))^c} d_X^2(x,x_0)d\boldsymbol{\gamma}(x,y) \\
&\leq \int\limits_{(B_R(x_0))^c} d_X^2(x,x_0)d\mu(x) + \int\limits_{X\times(B_R(y_0))^c} R^2 d\boldsymbol{\gamma}(x,y) \\
&\leq \int\limits_{(B_R(x_0))^c} d_X^2(x,x_0)d\mu(x) + \int\limits_{(B_R(y_0))^c} d_Y^2(y,y_0)d\nu(y),
\end{aligned}
$$

valid for any $\gamma \in \mathcal{A}dm(\mu,\nu)$ and the analogous one with the integral of $d_Y^2(y,y_0)$ in place of $d_X^2(x,x_0)$, show that if $\mathcal{K}_1 \subset \mathscr{P}_2(X)$ and $\mathcal{K}_2 \subset \mathscr{P}_2(Y)$ are 2-uniformly integrable, so is the set

$$\Big\{\gamma \in \mathscr{P}(X \times Y) : \pi^X_\#\gamma \in \mathcal{K}_1,\ \pi^Y_\#\gamma \in \mathcal{K}_2\Big\}.$$

■

We say that a function $f : X \to \mathbb{R}$ has quadratic growth provided

$$|f(x)| \leq a(d^2(x,x_0)+1), \tag{2.2}$$

for some $a \in \mathbb{R}$ and $x_0 \in X$. It is immediate to check that if f has quadratic growth and $\mu \in \mathscr{P}_2(X)$, then $f \in L^1(X,\mu)$.

The concept of 2-uniform integrability (in conjunction with tightness) in relation with convergence of integral of functions with quadratic growth, plays a role similar to the one played by tightness in relation with convergence of integral of bounded functions, as shown in the next proposition.

Proposition 2.4 *Let $(\mu_n) \subset \mathscr{P}_2(X)$ be a sequence narrowly converging to some μ. Then the following 3 properties are equivalent*

i) (μ_n) *is 2-uniformly integrable,*

ii) $\int f d\mu_n \to \int f d\mu$ *for any continuous f with quadratic growth,*

iii) $\int d^2(\cdot,x_0)d\mu_n \to \int d^2(\cdot,x_0)d\mu$ *for some $x_0 \in X$.*

Proof
(**i**) $\Rightarrow$ (**ii**). It is not restrictive to assume $f \geq 0$. Since any such f can be written as supremum of a family of continuous and bounded functions, it clearly holds

$$\int f d\mu \leq \liminf_{n\to\infty} \int f d\mu_n.$$

Thus we only have to prove the limsup inequality. Fix $\varepsilon > 0$, $x_0 \in X$ and find $R_\varepsilon > 1$ such that $\int_{X\setminus B_{R_\varepsilon}(x_0)} d^2(\cdot,x_0)d\mu_n \leq \varepsilon$ for every n. Now let χ be a function with bounded support, values in $[0,1]$ and identically 1 on B_{R_ε} and notice that for every $n \in \mathbb{N}$ it holds

$$\int f d\mu_n = \int f\chi d\mu_n + \int f(1-\chi)d\mu_n \leq \int f\chi d\mu_n + \int_{X\setminus B_{R_\varepsilon}} f d\mu_n \leq \int f\chi d\mu_n + 2a\varepsilon,$$

a being given by (2.2). Since $f\chi$ is continuous and bounded we have $\int f\chi d\mu_n \to \int f\chi d\mu$ and therefore

$$\varlimsup_{n\to\infty} \int f d\mu_n \leq \int f\chi d\mu + 2a\varepsilon \leq \int f d\mu + 2a\varepsilon.$$

Since $\varepsilon > 0$ was arbitrary, this part of the statement is proved.
(**ii**) $\Rightarrow$ (**iii**). Obvious.
(**iii**) $\Rightarrow$ (**i**). Argue by contradiction and assume that there exist $\varepsilon > 0$ and $\tilde{x}_0 \in X$ such that for every $R > 0$ it holds $\sup_{n\in\mathbb{N}} \int_{X\setminus B_R(\tilde{x}_0)} d^2(\cdot,\tilde{x}_0)d\mu_n > \varepsilon$. Then it is easy to see that it holds

$$\varlimsup_{n\to\infty} \int_{X\setminus B_R(x_0)} d^2(\cdot,x_0)d\mu_n > \varepsilon. \tag{2.3}$$

For every $R > 0$ let χ_R be a continuous cutoff function with values in $[0,1]$ supported on $B_R(x_0)$ and identically 1 on $B_{R/2}(x_0)$. Since $d^2(\cdot, x_0)\chi_R$ is continuous and bounded, we have

$$\begin{aligned}
\int d^2(\cdot, x_0)\chi_R d\mu &= \lim_{n\to\infty} \int d^2(\cdot, x_0)\chi_R d\mu_n \\
&= \lim_{n\to\infty} \left(\int d^2(\cdot, x_0) d\mu_n - \int d^2(\cdot, x_0)(1-\chi_R) d\mu_n \right) \\
&= \int d^2(\cdot, x_0) d\mu + \lim_{n\to\infty} - \int d^2(\cdot, x_0)(1-\chi_R) d\mu_n \\
&\leq \int d^2(\cdot, x_0) d\mu + \varliminf_{n\to\infty} - \int_{X\setminus B_R(x_0)} d^2(\cdot, x_0) d\mu_n \\
&= \int d^2(\cdot, x_0) d\mu - \varlimsup_{n\to\infty} \int_{X\setminus B_R(x_0)} d^2(\cdot, x_0) d\mu_n \\
&\leq \int d^2(\cdot, x_0) d\mu - \varepsilon,
\end{aligned}$$

having used (2.3) in the last step. Since

$$\int d^2(\cdot, x_0) d\mu = \sup_R \int d^2(\cdot, x_0)\chi_R d\mu \leq \int d^2(\cdot, x_0) d\mu - \varepsilon,$$

we got a contradiction. □

Proposition 2.5 (Stability of optimality) *The distance W_2 is lower semicontinuous w.r.t. narrow convergence of measures. Furthermore, if $(\boldsymbol{\gamma}_n) \subset \mathscr{P}_2(X^2)$ is a sequence of optimal plans which narrowly converges to $\boldsymbol{\gamma} \in \mathscr{P}_2(X^2)$, then $\boldsymbol{\gamma}$ is optimal as well.*

Proof Let (μ_n), $(\nu_n) \subset \mathscr{P}_2(X)$ be two sequences of measures narrowly converging to μ, $\nu \in \mathscr{P}_2(X)$ respectively. Pick $\boldsymbol{\gamma}_n \in \mathcal{O}pt(\mu_n, \nu_n)$ and use Remark 1.4 and Prokhorov theorem to get that $(\boldsymbol{\gamma}_n)$ admits a subsequence, not relabeled, narrowly converging to some $\boldsymbol{\gamma} \in \mathscr{P}(X^2)$. It is clear that $\pi^1_\#\boldsymbol{\gamma} = \mu$ and $\pi^2_\#\boldsymbol{\gamma} = \nu$, thus it holds

$$W_2^2(\mu, \nu) \leq \int d^2(x,y) d\boldsymbol{\gamma}(x,y) \leq \varliminf_{n\to\infty} \int d^2(x,y) d\boldsymbol{\gamma}_n(x,y) = \varliminf_{n\to\infty} W_2^2(\mu_n, \nu_n).$$

Now we pass to the second part of the statement, that is: we need to prove that with the same notation just used it holds $\boldsymbol{\gamma} \in \mathcal{O}pt(\mu, \nu)$. Choose $a(x) = b(x) = d^2(x, x_0)$ for some $x_0 \in X$ in the bound (1.4) and observe that since μ, $\nu \in \mathscr{P}_2(X)$ Theorem 1.13 applies, and thus optimality is equivalent to c-cyclical monotonicity of the support. The same for the plans $\boldsymbol{\gamma}_n$. Fix $N \in \mathbb{N}$ and pick $(x^i, y^i) \in \mathrm{supp}(\boldsymbol{\gamma})$, $i = 1, \ldots, N$. From the fact that $(\boldsymbol{\gamma}_n)$ narrowly converges to $\boldsymbol{\gamma}$ it is not hard to infer the existence of $(x^i_n, y^i_n) \in \mathrm{supp}(\boldsymbol{\gamma}_n)$ such that

$$\lim_{n\to\infty} \left(d(x^i_n, x^i) + d(y^i_n, y^i) \right) = 0, \qquad \forall i = 1, \ldots, N.$$

Thus the conclusion follows from the c-cyclical monotonicity of $\mathrm{supp}(\boldsymbol{\gamma}_n)$ and the continuity of the cost function. □

Now we are going to prove that $(\mathscr{P}_2(X), W_2)$ is a Polish space. In order to enable some constructions, we will use (a version of) Kolmogorov's theorem, which we recall without proof (see e.g. [31] §51).

Theorem 2.6 (Kolmogorov) *Let X be a Polish space and $\boldsymbol{\mu}_n \in \mathscr{P}(X^n)$, $n \in \mathbb{N}$, be a sequence of measures such that*

$$\pi_\#^{1,\ldots,n-1}\boldsymbol{\mu}_n = \boldsymbol{\mu}_{n-1}, \qquad \forall n \geq 2.$$

Then there exists a measure $\boldsymbol{\mu} \in X^{\mathbb{N}}$ such that

$$\pi_\#^{1,\ldots,n}\boldsymbol{\mu} = \boldsymbol{\mu}_n, \qquad \forall n \in \mathbb{N}.$$

Theorem 2.7 (Basic properties of the space $(\mathscr{P}_2(X), W_2)$) *Let (X, d) be complete and separable. Then*

$$W_2(\mu_n, \mu) \to 0 \qquad \Leftrightarrow \qquad \left\{ \begin{array}{rcll} \mu_n & \to & \mu & \textit{narrowly} \\ \displaystyle\int d^2(\cdot, x_0)d\mu_n & \to & \displaystyle\int d^2(\cdot, x_0)d\mu & \textit{for some } x_0 \in X. \end{array} \right. \tag{2.4}$$

Furthermore, the space $(\mathscr{P}_2(X), W_2)$ is complete and separable. Finally, $\mathcal{K} \subset \mathscr{P}_2(X)$ is relatively compact w.r.t. the topology induced by W_2 if and only if it is tight and 2-uniformly integrable.

Proof We start showing implication $\Rightarrow$ in (2.4). Thus assume that $W_2(\mu_n, \mu) \to 0$. Then

$$\left|\sqrt{\int d^2(\cdot, x_0)d\mu_n} - \sqrt{\int d^2(\cdot, x_0)d\mu}\right| = |W_2(\mu_n, \delta_{x_0}) - W_2(\mu, \delta_{x_0})| \leq W_2(\mu_n, \mu) \to 0.$$

To prove narrow convergence, for every $n \in \mathbb{N}$ choose $\boldsymbol{\gamma}_n \in \mathit{Opt}(\mu, \mu_n)$ and[2] use repeatedly the gluing lemma to find, for every $n \in \mathbb{N}$, a measure $\boldsymbol{\alpha}_n \in \mathscr{P}(X \times X^n)$ such that

$$\begin{aligned} \pi_\#^{0,n}\boldsymbol{\alpha}_n &= \boldsymbol{\gamma}_n, \\ \pi_\#^{0,1,\ldots,n-1}\boldsymbol{\alpha}_n &= \boldsymbol{\alpha}_{n-1}. \end{aligned}$$

Then by Kolmogorov's theorem we know that there exists a measure $\boldsymbol{\alpha} \in \mathscr{P}(X \times X^{\mathbb{N}})$ such that

$$\pi_\#^{0,1,\ldots,n}\boldsymbol{\alpha} = \boldsymbol{\alpha}_n, \qquad \forall n \in \mathbb{N}.$$

By construction we have

$$\|d(\pi^0, \pi^n)\|_{L^2(X \times X^{\mathbb{N}}, \boldsymbol{\alpha})} = \|d(\pi^0, \pi^n)\|_{L^2(X^2, \boldsymbol{\gamma}_n)} = W_2(\mu, \mu_n) \to 0.$$

Thus up to passing to a subsequence, not relabeled, we can assume that $\pi^n(\mathbf{x}) \to \pi^0(\mathbf{x})$ for $\boldsymbol{\alpha}$-almost any $\mathbf{x} \in X \times X^{\mathbb{N}}$. Now pick $f \in C_b(X)$ and use the dominated convergence theorem to get

$$\lim_{n\to\infty} \int f d\mu_n = \lim_{n\to\infty} \int f \circ \pi^n d\boldsymbol{\alpha} = \int f \circ \pi^0 d\boldsymbol{\alpha} = \int f d\mu.$$

[2] if closed balls in X are compact, the proof greatly simplifies. Indeed in this case the inequality $R^2\mu(X \setminus B_R(x_0)) \leq \int d^2(\cdot, x_0)d\mu$ and the uniform bound on the second moments yields that the sequence $n \mapsto \mu_n$ is tight. Thus to prove narrow convergence it is sufficient to check that $\int f d\mu_n \to \int f d\mu$ for every $f \in C_c(X)$. Since Lipschitz functions are dense in $C_c(X)$ w.r.t. uniform convergence, it is sufficient to check the convergence of the integral only for Lipschitz f's. This follows from the inequality

$$\begin{aligned} \left|\int f d\mu - \int f d\mu_n\right| &= \left|\int f(x) - f(y) d\boldsymbol{\gamma}_n(x, y)\right| \leq \int |f(x) - f(y)| d\boldsymbol{\gamma}_n(x, y) \\ &\leq \mathrm{Lip}(f) \int d(x, y) d\boldsymbol{\gamma}_n(x, y) \leq \mathrm{Lip}(f)\sqrt{\int d^2(x, y) d\boldsymbol{\gamma}_n(x, y)} = \mathrm{Lip}(f) W_2(\mu, \mu_n). \end{aligned}$$

Since the argument does not depend on the subsequence chosen, the claim is proved.

We pass to the converse implication in (2.4). Pick $\boldsymbol{\gamma}_n \in Opt(\mu, \mu_n)$ and use Remark 1.4 to get that the sequence $(\boldsymbol{\gamma}_n)$ is tight, hence, up to passing to a subsequence, we can assume that it narrowly converges to some $\boldsymbol{\gamma}$. By Proposition 2.5 we know that $\boldsymbol{\gamma} \in Opt(\mu, \mu)$, which forces $\int d^2(x,y)d\boldsymbol{\gamma}(x,y) = 0$. By Proposition 2.4 and our assumption on $(\mu_n), \mu$ we know that (μ_n) is 2-uniformly integrable, thus by Remark 2.3 again we know that $(\boldsymbol{\gamma}_n)$ is 2-uniformly integrable as well. Since the map $(x,y) \mapsto d^2(x,y)$ has quadratic growth in X^2 it holds

$$\lim_{n\to\infty} W_2^2(\mu_n, \mu) = \lim_{n\to\infty} \int d^2(x,y)d\boldsymbol{\gamma}_n(x,y) = \int d^2(x,y)d\boldsymbol{\gamma}(x,y) = 0.$$

Now we prove that $(\mathscr{P}_2(X), W_2)$ is complete. Pick a Cauchy sequence (μ_n) and assume[3], without loss of generality, that $\sum_n W_2(\mu_n, \mu_{n+1}) < \infty$. For every $n \in \mathbb{N}$ choose $\boldsymbol{\gamma}_n \in Opt(\mu_n, \mu_{n+1})$ and use repeatedly the gluing lemma to find, for every $n \in \mathbb{N}$, a measure $\boldsymbol{\beta}_n \in \mathscr{P}_2(X^n)$ such that

$$\begin{aligned} \pi^{n,n+1}_{\#}\boldsymbol{\beta}_n &= \boldsymbol{\gamma}_n, \\ \pi^{1,\ldots,n-1}_{\#}\boldsymbol{\beta}_n &= \boldsymbol{\alpha}_{n-1} \end{aligned}$$

By Kolmogorov's theorem we get the existence of a measure $\boldsymbol{\beta} \in \mathscr{P}(X^{\mathbb{N}})$ such that $\pi^{1,\ldots,n}_{\#}\boldsymbol{\beta} = \boldsymbol{\beta}_n$ for every $n \in \mathbb{N}$. The inequality

$$\sum_{n=1}^{\infty} \|d(\pi^i, \pi^{i+1})\|_{L^2(X^{\mathbb{N}},\boldsymbol{\beta})} = \sum_{n=1}^{\infty} \|d(\pi^i, \pi^{i+1})\|_{L^2(X^2,\boldsymbol{\gamma}_i)} = \sum_{n=1}^{\infty} W_2(\mu_i, \mu_{i+1}) < \infty,$$

shows that $n \mapsto \pi^n : X^{\mathbb{N}} \to X$ is a Cauchy sequence in $L^2(\boldsymbol{\beta}, X)$, i.e. the space of maps $f : X^{\mathbb{N}} \to X$ such that $\int d^2(f(y), x_0)d\boldsymbol{\beta}(y) < \infty$ for some, and thus every, $x_0 \in X$ endowed with the distance $\tilde{d}(f,g) := \sqrt{\int d^2(f(y), g(y))d\boldsymbol{\beta}(y)}$. Since X is complete, $L^2(\boldsymbol{\beta}, X)$ is complete as well, and therefore there exists a limit map π^∞ of the Cauchy sequence (π^n). Define $\mu := \pi^\infty_{\#}\boldsymbol{\beta}$ and notice that by (2.1) we have

$$W_2^2(\mu, \mu_n) \leq \int d^2(\pi^\infty, \pi^n)d\boldsymbol{\beta} \to 0,$$

so that μ is the limit of the Cauchy sequence (μ_n) in $(\mathscr{P}_2(X), W_2)$. The fact that $(\mathscr{P}_2(X), W_2)$ is separable follows from (2.4) by considering the set of finite convex combinations of Dirac masses centered at points in a dense countable set in X with rational coefficients. The last claim now follows. $\square$

Remark 2.8 (On compactness properties of $\mathscr{P}_2(X)$) An immediate consequence of the above theorem is the fact that if X is compact, then $(\mathscr{P}_2(X), W_2)$ is compact as well: indeed, in this case the equivalence (2.4) tells that convergence in $\mathscr{P}_2(X)$ is equivalent to weak convergence.

It is also interesting to notice that if X is unbounded, then $\mathscr{P}_2(X)$ is not locally compact. Actually, for any measure $\mu \in \mathscr{P}_2(X)$ and any $r > 0$, the closed ball of radius r around μ is not compact. To see this, fix $\overline{x} \in X$ and find a sequence $(x_n) \subset X$ such that $d(x_n, \overline{x}) \to \infty$. Now define the

[3] again, if closed balls in X are compact the argument simplifies. Indeed from the uniform bound on the second moments and the inequality $R^2\mu(X \setminus B_R(x_0)) \leq \int_{X\setminus B_R(x_0)} d^2(\cdot, x_0)d\mu$ we get the tightness of the sequence. Hence up to pass to a subsequence we can assume that (μ_n) narrowly converges to a limit measure μ, and then using the lower semicontinuity of W_2 w.r.t. narrow convergence we can conclude $\overline{\lim}_n W_2(\mu, \mu_n) \leq \overline{\lim}_n \underline{\lim}_m W_2(\mu_m, \mu_n) = 0$

measures $\mu_n := (1-\varepsilon_n)\mu + \varepsilon_n \delta_{x_n}$, where ε_n is chosen such that $\varepsilon_n d^2(\overline{x}, x_n) = r^2$. To bound from above $W_2^2(\mu, \mu_n)$, leave fixed $(1-\varepsilon_n)\mu$, move $\varepsilon_n \mu$ to $\overline{x}$ and then move $\varepsilon_n \delta_{\overline{x}}$ into $\varepsilon_n \delta_{x_n}$, this gives

$$W_2^2(\mu, \mu_n) \leq \varepsilon_n \left(\int d^2(x, \overline{x}) d\mu(x) + d^2(x_n, \overline{x}) \right),$$

so that $\overline{\lim} W_2(\mu, \mu_n) \leq r$. Conclude observing that

$$\underline{\lim}_{n\to\infty} \int d^2(x, \overline{x}) d\mu_n = \underline{\lim}_{n\to\infty} (1-\varepsilon_n) \int d^2(x, \overline{x}) d\mu + \varepsilon_n d^2(x_n, \overline{x}) = \int d^2(x, \overline{x}) d\mu + r^2,$$

thus the second moments do not converge. Since clearly (μ_n) weakly converges to μ, we proved that there is no local compactness. ■

2.2 X geodesic space

In this section we prove that if the base space (X, d) is geodesic, then the same is true also for $(\mathscr{P}_2(X), W_2)$ and we will analyze the properties of this latter space.

Let us recall that a curve $\gamma : [0,1] \to X$ is called *constant speed geodesic* provided

$$d(\gamma_t, \gamma_s) = |t-s| d(\gamma_0, \gamma_1), \qquad \forall t, s \in [0,1], \tag{2.5}$$

or equivalently if $\leq$ always holds.

Definition 2.9 (Geodesic space) *A metric space (X, d) is called* geodesic *if for every x, $y \in X$ there exists a constant speed geodesic connecting them, i.e. a constant speed geodesic such that $\gamma_0 = x$ and $\gamma_1 = y$.*

Before entering into the details, let us describe an important example. Recall that $X \ni x \mapsto \delta_x \in \mathscr{P}_2(X)$ is an isometry. Therefore if $t \mapsto \gamma_t$ is a constant speed geodesic on X connecting x to y, the curve $t \mapsto \delta_{\gamma_t}$ is a constant speed geodesic on $\mathscr{P}_2(X)$ which connects δ_x to δ_y. The important thing to notice here is that the natural way to interpolate between δ_x and δ_y is given by this - so called - *displacement interpolation*. Conversely, observe that the classical linear interpolation

$$t \mapsto \mu_t := (1-t)\delta_x + t\delta_y,$$

produces a curve which has infinite length as soon as $x \neq y$ (because $W_2(\mu_t, \mu_s) = \sqrt{|t-s|} d(x, y)$), and thus is unnatural in this setting.

We will denote by $\mathrm{Geod}(X)$ the metric space of all constant speed geodesics on X endowed with the $\sup$ norm. With some work it is possible to show that $\mathrm{Geod}(X)$ is complete and separable as soon as X is (we omit the details). The *evaluation maps* $\mathrm{e}_t : \mathrm{Geod}(X) \to X$ are defined for every $t \in [0,1]$ by

$$\mathrm{e}_t(\gamma) := \gamma_t. \tag{2.6}$$

Theorem 2.10 *Let (X, d) be Polish and geodesic. Then $(\mathscr{P}_2(X), W_2)$ is geodesic as well. Furthermore, the following two are equivalent:*

i) *$t \mapsto \mu_t \in \mathscr{P}_2(X)$ is a constant speed geodesic,*

ii) *There exists a measure $\boldsymbol{\mu} \in \mathscr{P}_2(\mathrm{Geod}(X))$ such that $(\mathrm{e}_0, \mathrm{e}_1)_\# \boldsymbol{\mu} \in \mathit{Opt}(\mu_0, \mu_1)$ and*

$$\mu_t = (\mathrm{e}_t)_\# \boldsymbol{\mu}. \tag{2.7}$$

Proof Choose μ^0, $\mu^1 \in \mathscr{P}_2(X)$ and find an optimal plan $\boldsymbol{\gamma} \in \mathit{Opt}(\mu,\nu)$. By Lemma 2.11 below and classical measurable selection theorems we know that there exists a Borel map $\mathrm{GeodSel}: X^2 \to \mathrm{Geod}(X)$ such that for any x, $y \in X$ the curve $\mathrm{GeodSel}(x,y)$ is a constant speed geodesic connecting x to y. Define the Borel probability measure $\boldsymbol{\mu} \in \mathscr{P}(\mathrm{Geod}(X))$ by

$$\boldsymbol{\mu} := \mathrm{GeodSel}_\# \boldsymbol{\gamma},$$

and the measures $\mu_t \in \mathscr{P}(X)$ by $\mu_t := (\mathrm{e}_t)_\# \boldsymbol{\mu}$.

We claim that $t \mapsto \mu_t$ is a constant speed geodesic connecting μ^0 to μ^1. Consider indeed the map $(\mathrm{e}_0, \mathrm{e}_1): \mathrm{Geod}(X) \to X^2$ and observe that from $(\mathrm{e}_0,\mathrm{e}_1)\Big(\mathrm{GeodSel}(x,y)\Big) = (x,y)$ we get

$$(\mathrm{e}_0, \mathrm{e}_1)_\# \boldsymbol{\mu} = \boldsymbol{\gamma}. \tag{2.8}$$

In particular, $\mu_0 = (\mathrm{e}_0)_\# \boldsymbol{\mu} = \pi^1_\# \boldsymbol{\gamma} = \mu^0$, and similarly $\mu_1 = \mu^1$, so that the curve $t \mapsto \mu_t$ connects μ^0 to μ^1. The facts that the measures μ_t have finite second moments and (μ_t) is a constant speed geodesic follow from

$$\begin{aligned} W_2^2(\mu_t, \mu_s) &\overset{(2.7),(2.1)}{\leq} \int d^2(\mathrm{e}_t(\gamma), \mathrm{e}_s(\gamma))\, d\boldsymbol{\mu}(\gamma) \\ &\overset{(2.5)}{=} (t-s)^2 \int d^2(\mathrm{e}_0(\gamma), \mathrm{e}_1(\gamma))\, d\boldsymbol{\mu}(\gamma) \\ &\overset{(2.8)}{=} (t-s)^2 \int d^2(x,y)\, d\boldsymbol{\gamma}(x,y) = (t-s)^2 W_2^2(\mu^0,\mu^1). \end{aligned}$$

The fact that (ii) implies (i) follows from the same kind of argument just used. So, we turn to $(i) \Rightarrow (ii)$. For $n \geq 0$ we use iteratively the gluing Lemma 2.1 and the Borel map GeodSel to build a measure $\boldsymbol{\mu}^n \in \mathscr{P}(C([0,1],X))$ such that

$$\big(\mathrm{e}_{i/2^n}, \mathrm{e}_{(i+1)/2^n}\big)_\# \boldsymbol{\mu}^n \in \mathit{Opt}(\mu_{i/2^n}, \mu_{(i+1)/2^n}), \qquad \forall i = 0, \ldots, 2^n - 1,$$

and $\boldsymbol{\mu}^n$-a.e. γ is a geodesic in the intervals $[i/2^n, (i+1)/2^n]$, $i = 0, \ldots, 2^n-1$. Fix n and observe that for any $0 \leq j < k \leq 2^n$ it holds

$$\begin{aligned} \big\| d\big(\mathrm{e}_{j/2^n}, \mathrm{e}_{k/2^n}\big) \big\|_{L^2(\boldsymbol{\mu}^n)} &\leq \Big\| \sum_{i=j}^{k-1} d\big(\mathrm{e}_{i/2^n}, \mathrm{e}_{(i+1)/2^n}\big) \Big\|_{L^2(\boldsymbol{\mu}^n)} \leq \sum_{i=j}^{k-1} \big\| d\big(\mathrm{e}_{i/2^n}, \mathrm{e}_{(i+1)/2^n}\big) \big\|_{L^2(\boldsymbol{\mu}^n)} \\ &= \sum_{i=j}^{k-1} W_2(\mu_{i/2^n}, \mu_{(i+1)/2^n}) = W_2(\mu_{j/2^n}, \mu_{k/2^n}). \end{aligned} \tag{2.9}$$

Therefore it holds

$$\big(\mathrm{e}_{j/2^n}, \mathrm{e}_{k/2^n}\big)_\# \boldsymbol{\mu}^n \in \mathit{Opt}(\mu_{j/2^n}, \mu_{k/2^n}), \qquad \forall j,k \in \{0, \ldots, 2^n\}.$$

Also, since the inequalities in (2.9) are equalities, it is not hard to see that for $\boldsymbol{\mu}^n$-a.e. γ the points $\gamma_{i/2^n}$, $i = 0, \ldots, 2^n$, must lie along a geodesic and satisfy $d(\gamma_{i/2^n}, \gamma_{(i+1)/2^n}) = d(\gamma_0, \gamma_1)/2^n$, $i = 0, \ldots, 2^n - 1$. Hence $\boldsymbol{\mu}^n$-a.e. γ is a constant speed geodesic and thus $\boldsymbol{\mu}^n \in \mathscr{P}(\mathrm{Geod}(X))$. Now suppose for a moment that $(\boldsymbol{\mu}^n)$ narrowly converges - up to pass to a subsequence - to some $\boldsymbol{\mu} \in \mathscr{P}(\mathrm{Geod}(X))$. Then the continuity of the evaluation maps e_t yields that for any $t \in [0,1]$ the sequence $n \mapsto (\mathrm{e}_t)_\# \boldsymbol{\mu}^n$ narrowly converges to $(\mathrm{e}_t)_\# \boldsymbol{\mu}$ and this, together with the uniform bound (2.9), easily implies that $\boldsymbol{\mu}$ satisfies (2.7).

Thus to conclude it is sufficient to show that some subsequence of $(\boldsymbol{\mu}_n)$ has a narrow limit[4]. We will prove this by showing that $\boldsymbol{\mu}^n \in \mathscr{P}_2(\mathrm{Geod}(X))$ for every $n \in \mathbb{N}$ and that some subsequence is a Cauchy sequence in $(\mathscr{P}_2(\mathrm{Geod}(X)), \mathbf{W}_2)$, $\mathbf{W}_2$ being the Wasserstein distance built over $\mathrm{Geod}(X)$ endowed with the $\sup$ distance, so that by Theorem 2.7 we conclude.

We know by Remark 1.4, Remark 2.3 and Theorem 2.7 that for every $n \in \mathbb{N}$ the set of plans $\boldsymbol{\alpha} \in \mathscr{P}_2(X^{2^n+1})$ such that $\pi^i_\# \boldsymbol{\alpha} = \mu_{i/2^n}$ for $i = 0, \ldots, 2^n$, is compact in $\mathscr{P}_2(X^{2^n+1})$. Therefore a diagonal argument tells that possibly passing to a subsequence, not relabeled, we may assume that for every $n \in \mathbb{N}$ the sequence

$$m \mapsto \prod_{i=0}^{2^n} (\mathrm{e}_{i/2^n})_\# \boldsymbol{\mu}^m$$

converges to some plan w.r.t. the distance W_2 on X^{2^n+1}.

Now fix $n \in \mathbb{N}$ and notice that for $t \in [i/2^n, (i+1)/2^n]$ and $\gamma, \tilde{\gamma} \in \mathrm{Geod}(X)$ it holds

$$d\big(\gamma_t, \tilde{\gamma}_t\big) \leq d\big(\gamma_{i/2^n}, \tilde{\gamma}_{(i+1)/2^n}\big) + \frac{1}{2^n}\Big(d(\gamma_0, \gamma_1) + d(\tilde{\gamma}_0, \tilde{\gamma}_1)\Big),$$

and therefore squaring and then taking the $\sup$ over $t \in [0,1]$ we get

$$\sup_{t \in [0,1]} d^2(\gamma_t, \tilde{\gamma}_t) \leq 2 \sum_{i=0}^{2^n-1} d^2\big(\gamma_{i/2^n}, \tilde{\gamma}_{(i+1)/2^n}\big) + \frac{1}{2^{n-2}}\Big(d^2(\gamma_0, \gamma_1) + d^2(\tilde{\gamma}_0, \tilde{\gamma}_1)\Big). \tag{2.10}$$

Choosing $\tilde{\gamma}$ to be a constant geodesic and using (2.9), we get that $\boldsymbol{\mu}^m \in \mathscr{P}_2(\mathrm{Geod}(X))$ for every $m \in \mathbb{N}$. Now, for any given $\boldsymbol{\nu}, \tilde{\boldsymbol{\nu}} \in \mathscr{P}(\mathrm{Geod}(X))$, by a gluing argument (Lemma 2.12 below with $\boldsymbol{\nu}, \tilde{\boldsymbol{\nu}}$ in place of $\nu, \tilde{\nu}$, $Y = \mathrm{Geod}(X)$, $Z = X^{2^n+1}$) we can find a plan $\boldsymbol{\beta} \in \mathscr{P}([\mathrm{Geod}(X)]^2)$ such that

$$\begin{aligned} \pi^1_\# \boldsymbol{\beta} &= \boldsymbol{\nu}, \\ \pi^2_\# \boldsymbol{\beta} &= \tilde{\boldsymbol{\nu}}, \\ \Big(\big(\mathrm{e}_0, \ldots, \mathrm{e}_{i/2^n}, \ldots, \mathrm{e}_1\big) \circ \pi^1, \big(\mathrm{e}_0, \ldots, \mathrm{e}_{i/2^n}, \ldots, \mathrm{e}_1\big) \circ \pi^2\Big)_\# \boldsymbol{\beta} &\in \mathit{Opt}\Big(\prod_{i=0}^{2^n} (\mathrm{e}_{i/2^n})_\# \boldsymbol{\nu}, \prod_{i=0}^{2^n} (\mathrm{e}_{i/2^n})_\# \tilde{\boldsymbol{\nu}}\Big), \end{aligned}$$

where optimality between $\prod_{i=0}^{2^n} (\mathrm{e}_{i/2^n})_\# \boldsymbol{\nu}$ and $\prod_{i=0}^{2^n} (\mathrm{e}_{i/2^n})_\# \tilde{\boldsymbol{\nu}}$ is meant w.r.t. the Wasserstein distance on $\mathscr{P}_2(X^{2^n+1})$. Using $\boldsymbol{\beta}$ to bound from above $\mathbf{W}_2(\boldsymbol{\nu}, \tilde{\boldsymbol{\nu}})$ and using (2.10) we get that for every couple of measures $\boldsymbol{\nu}$, $\tilde{\boldsymbol{\nu}} \in \mathscr{P}_2(\mathrm{Geod}(X))$ it holds

$$\begin{aligned} \mathbf{W}_2^2(\boldsymbol{\nu}, \tilde{\boldsymbol{\nu}}) \ &\leq 2 W_2^2\Big(\prod_{i=0}^{2^n} (\mathrm{e}_{i/2^n})_\# \boldsymbol{\nu}, \prod_{i=0}^{2^n} (\mathrm{e}_{i/2^n})_\# \tilde{\boldsymbol{\nu}}\Big) \\ &\quad + \frac{1}{2^{n-2}}\left(\int d^2(\gamma_0, \gamma_1) d\boldsymbol{\nu}(\gamma) + \int d^2(\tilde{\gamma}_0, \tilde{\gamma}_1) d\boldsymbol{\nu}(\tilde{\gamma})\right) \end{aligned}$$

[4]as for Theorem 2.7 everything is simpler if closed balls in X are compact. Indeed, observe that a geodesic connecting two points in $B_R(x_0)$ lies entirely on the compact set $\overline{B_{2R}(x_0)}$, and that the set of geodesics lying on a given compact set is itself compact in $\mathrm{Geod}(X)$, so that the tightness of $(\boldsymbol{\mu}^n)$ follows directly from the one of $\{\mu_0, \mu_1\}$.

Plugging $\boldsymbol{\nu} = \boldsymbol{\mu}^m$, $\tilde{\boldsymbol{\nu}} = \boldsymbol{\mu}^{m'}$ and recalling that $W_2\Big(\prod_{i=0}^{2^n}(\mathrm{e}_{i/2^n})_\#\boldsymbol{\mu}^m, \prod_{i=0}^{2^n}(\mathrm{e}_{i/2^n})_\#\boldsymbol{\mu}^{m'}\Big) \to 0$ as $m,\, m' \to +\infty$ for every $n \in \mathbb{N}$ we get that

$$\overline{\lim_{m,\, m' \to \infty}} \mathbf{W}_2^2(\boldsymbol{\mu}^m, \boldsymbol{\mu}^{m'}) \le \frac{1}{2^{n-2}}\left(\int d^2(\gamma_0,\gamma_1)\,d\boldsymbol{\mu}^m(\gamma) + \int d^2(\tilde{\gamma}_0,\tilde{\gamma}_1)\,d\boldsymbol{\mu}^{m'}(\tilde{\gamma})\right)$$
$$= \frac{1}{2^{n-3}} W_2^2(\mu_0,\mu_1).$$

Letting $n \to \infty$ we get that $(\boldsymbol{\mu}^m) \subset \mathscr{P}_2(\mathrm{Geod}(X))$ is a Cauchy sequence and the conclusion. $\square$

Lemma 2.11 *The multivalued map from $G : X^2 \to \mathrm{Geod}(X)$ which associates to each pair (x, y) the set $G(x, y)$ of constant speed geodesics connecting x to y has closed graph.*

Proof Straightforward. $\square$

Lemma 2.12 (A variant of gluing) *Let Y, Z be Polish spaces, $\nu, \tilde{\nu} \in \mathscr{P}(Y)$ and $f, g : Y \to Z$ be two Borel maps. Let $\boldsymbol{\gamma} \in \mathcal{A}dm(f_\#\nu, g_\#\tilde{\nu})$. Then there exists a plan $\boldsymbol{\beta} \in \mathscr{P}(Y^2)$ such that*

$$\pi^1_\#\boldsymbol{\beta} = \nu,$$
$$\pi^2_\#\boldsymbol{\beta} = \tilde{\nu},$$
$$(f \circ \pi^1, g \circ \pi^2)_\#\boldsymbol{\beta} = \boldsymbol{\gamma}.$$

Proof Let $\{\nu_z\}, \{\tilde{\nu}_{\tilde{z}}\}$ be the disintegrations of $\nu, \tilde{\nu}$ w.r.t. f, g respectively. Then define

$$\boldsymbol{\beta} := \int_{Z^2} \nu_z \times \tilde{\nu}_{\tilde{z}}\, d\boldsymbol{\gamma}(z, \tilde{z}).$$

$\square$

Remark 2.13 (The Hilbert case) If X is an Hilbert space, then for every $x,\, y \in X$ there exists only one constant speed geodesic connecting them: the curve $t \mapsto (1-t)x + ty$. Thus Theorem 2.10 reads as: $t \mapsto \mu_t$ is a constant speed geodesic if and only if there exists an optimal plan $\boldsymbol{\gamma} \in \mathcal{O}pt(\mu_0, \mu_1)$ such that

$$\mu_t = \big((1-t)\pi^1 + t\pi^2\big)_\#\boldsymbol{\gamma}.$$

If $\boldsymbol{\gamma}$ is induced by a map T, the formula further simplifies to

$$\mu_t = \big((1-t)Id + tT\big)_\#\mu_0. \tag{2.11}$$

$\blacksquare$

Remark 2.14 A slight modification of the arguments presented in the second part of the proof of Theorem 2.10 shows that if (X, d) is Polish and $(\mathscr{P}_2(X), W_2)$ is geodesic, then (X, d) is geodesic as well. Indeed, given $x, y \in X$ and a geodesic (μ_t) connecting δ_x to δ_y, we can build a measure $\boldsymbol{\mu} \in \mathscr{P}(\mathrm{Geod}(X))$ satisfying (2.7). Then every $\gamma \in \mathrm{supp}(\boldsymbol{\mu})$ is a geodesic connecting x to y. $\blacksquare$

Definition 2.15 (Non branching spaces) *A geodesic space (X, d) is said non branching if for any $t \in (0, 1)$ a constant speed geodesic γ is uniquely determined by its initial point γ_0 and by the point γ_t. In other words, (X, d) is non branching if the map*

$$\mathrm{Geod}(X) \ni \gamma \mapsto (\gamma_0, \gamma_t) \in X^2,$$

is injective for any $t \in (0, 1)$.

Non-branching spaces are interesting from the optimal transport point of view, because for such spaces the behavior of geodesics in $\mathscr{P}_2(X)$ is particularly nice: optimal transport plan from intermediate measures to other measures along the geodesic are unique and induced by maps (it is quite surprising that such a statement is true in this generality - compare the assumption of the proposition below with the ones of Theorems 1.26, 1.33). Examples of non-branching spaces are Riemannian manifolds, Banach spaces with strictly convex norms and Alexandrov spaces with curvature bounded below. Examples of branching spaces are Banach spaces with non strictly convex norms.

Proposition 2.16 (Non branching and interior regularity) *Let (X, d) be a Polish, geodesic, non branching space. Then $(\mathscr{P}_2(X), W_2)$ is non branching as well. Furthermore, if $(\mu_t) \subset \mathscr{P}_2(X)$ is a constant speed geodesic, then for every $t \in (0,1)$ there exists only one optimal plan in $\mathit{Opt}(\mu_0, \mu_t)$ and this plan is induced by a map from μ_t. Finally, the measure $\boldsymbol{\mu} \in \mathscr{P}(\mathrm{Geod}(X))$ associated to (μ_t) via* (2.7) *is unique.*

Proof Let $(\mu_t) \subset \mathscr{P}_2(X)$ be a constant speed geodesic and fix $t_0 \in (0,1)$. Pick $\boldsymbol{\gamma}^1 \in \mathit{Opt}(\mu_0, \mu_{t_0})$ and $\boldsymbol{\gamma}^2 \in \mathit{Opt}(\mu_{t_0}, \mu_1)$. We want to prove that both $\boldsymbol{\gamma}^1$ and $\boldsymbol{\gamma}^2$ are induced by maps from μ_{t_0}. To this aim use the gluing lemma to find a 3-plan $\boldsymbol{\alpha} \in \mathscr{P}_2(X^3)$ such that

$$\pi^{1,2}_{\#}\boldsymbol{\alpha} = \boldsymbol{\gamma}^1,$$
$$\pi^{2,3}_{\#}\boldsymbol{\alpha} = \boldsymbol{\gamma}^2,$$

and observe that since (μ_t) is a geodesic it holds

$$\begin{aligned}\|d(\pi^1, \pi^3)\|_{L^2(\boldsymbol{\alpha})} &\le \|d(\pi^1, \pi^2) + d(\pi^2, \pi^3)\|_{L^2(\boldsymbol{\alpha})} \le \|d(\pi^1, \pi^2)\|_{L^2(\boldsymbol{\alpha})} + \|d(\pi^2, \pi^3)\|_{L^2(\boldsymbol{\alpha})} \\ &= \|d(\pi^1, \pi^2)\|_{L^2(\boldsymbol{\gamma}^1)} + \|d(\pi^1, \pi^2)\|_{L^2(\boldsymbol{\gamma}^2)} = W_2(\mu_0, \mu_{t_0}) + W_2(\mu_{t_0}, \mu_1) \\ &= W_2(\mu_0, \mu_1),\end{aligned}$$

so that $(\pi^1, \pi^3)_{\#}\boldsymbol{\alpha} \in \mathit{Opt}(\mu_0, \mu_1)$. Also, since the first inequality is actually an equality, we have that $d(x, y) + d(y, z) = d(x, z)$ for $\boldsymbol{\alpha}$-a.e. (x, y, z), which means that x, y, z lie along a geodesic. Furthermore, since the second inequality is an equality, the functions $(x, y, z) \mapsto d(x, y)$ and $(x, y, z) \mapsto d(y, z)$ are each a positive multiple of the other in $\mathrm{supp}(\boldsymbol{\alpha})$. It is then immediate to verify that for every $(x, y, z) \in \mathrm{supp}(\boldsymbol{\alpha})$ it holds

$$d(x, y) = (1 - t_0)d(x, z),$$
$$d(y, z) = t_0 d(x, z).$$

We now claim that for $(x, y, z), (x', y', z') \in \mathrm{supp}(\boldsymbol{\alpha})$ it holds $(x, y, z) = (x', y', z')$ if and only if $y = y'$. Indeed, pick (x, y, z), $(x', y, z') \in \mathrm{supp}(\boldsymbol{\alpha})$ and assume, for instance, that $z \neq z'$. Since $(\pi^1, \pi^3)_{\#}\boldsymbol{\alpha}$ is an optimal plan, by the cyclical monotonicity of its support we know that

$$\begin{aligned}d^2(x, z) + d^2(x', z') &\le d^2(x, z') + d^2(x', z) \le \big(d(x, y) + d(y, z')\big)^2 + \big(d(x', y) + d(y, z)\big)^2 \\ &= \big((1 - t_0)d(x, z) + t_0 d(x', z')\big)^2 + \big((1 - t_0)d(x', z') + t_0 d(x, z)\big)^2,\end{aligned}$$

which, after some manipulation, gives $d(x, z) = d(x', z') =: D$. Again from the cyclical monotonicity of the support we have $2D^2 \le d^2(x, z') + d^2(x', z)$, thus either $d(x', z)$ or $d(x, z')$ is $\ge$ than D. Say $d(x, z') \ge D$, so that it holds

$$D \le d(x, z') \le d(x, y) + d(y, z') = (1 - t_0)D + t_0 D = D,$$

which means that the triple of points (x, y, z') lies along a geodesic. Since (x, y, z) lies on a geodesic as well, by the non-branching hypothesis we get a contradiction.

Thus the map $\mathrm{supp}(\boldsymbol{\alpha}) \ni (x, y, z) \mapsto y$ is injective. This means that there exists two maps $f, g : X \to X$ such that $(x, y, z) \in \mathrm{supp}(\boldsymbol{\alpha})$ if and only if $x = f(y)$ and $z = g(y)$. This is the same as to say that $\boldsymbol{\gamma}^1$ is induced by f and $\boldsymbol{\gamma}^2$ is induced by g.

To summarize, we proved that given $t_0 \in (0, 1)$, every optimal plan $\boldsymbol{\gamma} \in \mathit{Opt}(\mu_0, \mu_{t_0})$ is induced by a map from μ_{t_0}. Now we claim that the optimal plan is actually unique. Indeed, if there are two of them induced by two different maps, say f and f', then the plan

$$\frac{1}{2}\big((f, Id)_{\#}\mu_{\mu_{t_0}} + (f', Id)_{\#}\mu_{\mu_{t_0}}\big),$$

would be optimal and not induced by a map.

It remains to prove that $\mathscr{P}_2(X)$ is non branching. Choose $\boldsymbol{\mu} \in \mathscr{P}_2(\mathrm{Geod}(X))$ such that (2.7) holds, fix $t_0 \in (0, 1)$ and let $\boldsymbol{\gamma}$ be the unique optimal plan in $\mathit{Opt}(\mu_0, \mu_{t_0})$. The thesis will be proved if we show that $\boldsymbol{\mu}$ depends only on $\boldsymbol{\gamma}$. Observe that from Theorem 2.10 and its proof we know that

$$(\mathrm{e}_0, \mathrm{e}_{t_0})_{\#}\boldsymbol{\mu} \in \mathit{Opt}(\mu_0, \mu_{t_0}),$$

and thus $(\mathrm{e}_0, \mathrm{e}_{t_0})_{\#}\boldsymbol{\mu} = \boldsymbol{\gamma}$. By the non-branching hypothesis we know that $(\mathrm{e}_0, \mathrm{e}_{t_0}) : \mathrm{Geod}(X) \to X^2$ is injective. Thus it it invertible on its image: letting F the inverse map, we get

$$\boldsymbol{\mu} = F_{\#}\boldsymbol{\gamma},$$

and the thesis is proved. □

Theorem 2.10 tells us not only that geodesics exists, but provides also a natural way to "interpolate" optimal plans: once we have the measure $\boldsymbol{\mu} \in \mathscr{P}(\mathrm{Geod}(X))$ satisfying (2.7), an optimal plan from μ_t to μ_s is simply given by $(\mathrm{e}_t, \mathrm{e}_s)_{\#}\boldsymbol{\mu}$. Now, we know that the transport problem has a natural dual problem, which is solved by the Kantorovich potential. It is then natural to ask how to interpolate potentials. In other words, if (φ, φ^{c+}) are $c-$conjugate Kantorovich potentials for (μ_0, μ_1), is there a simple way to find out a couple of Kantorovich potentials associated to the couple μ_t, μ_s? The answer is yes, and it is given - shortly said - by the solution of an Hamilton-Jacobi equation. To see this, we first define the *Hopf-Lax* evolution semigroup H_t^s (which in $\mathbb{R}^d$ produces the viscosity solution of the Hamilton-Jacobi equation) via the following formula:

$$H_t^s(\psi)(x) := \begin{cases} \inf\limits_{y \in X} \dfrac{d^2(x, y)}{s - t} + \psi(y), & \text{if } t < s, \\[2ex] \psi(x), & \text{if } t = s, \\[2ex] \sup\limits_{y \in X} -\dfrac{d^2(x, y)}{s - t} + \psi(y), & \text{if } t > s, \end{cases} \tag{2.12}$$

To fully appreciate the mechanisms behind the theory, it is better to introduce the *rescaled costs* $c^{t,s}$ defined by

$$c^{t,s}(x, y) := \frac{d^2(x, y)}{s - t}, \qquad \forall t < s,\ x, y \in X.$$

Observe that for $t < r < s$

$$c^{t,r}(x, y) + c^{r,s}(y, z) \geq c^{t,s}(x, z), \qquad \forall x, y, z \in X,$$

and equality holds if and only if there is a constant speed geodesic $\gamma : [t, s] \to X$ such that $x = \gamma_t$, $y = \gamma_r$, $z = \gamma_s$. The notions of $c_+^{t,s}$ and $c_-^{t,s}$ transforms, convexity/concavity and sub/super-differential are defined as in Section 1.2, Definitions 1.8, 1.9 and 1.10.

The basic properties of the Hopf-Lax formula are collected in the following proposition:

Proposition 2.17 (Basic properties of the Hopf-Lax formula) *We have the following three properties:*

(i) For any $t,\ s \in [0,1]$ the map H_t^s is order preserving, that is $\phi \le \psi \Rightarrow H_t^s(\phi) \le H_t^s(\psi)$.

(ii) For any $t < s \in [0,1]$ it holds

$$H_s^t\big(H_t^s(\phi)\big) = \phi^{c_-^{t,s} c_-^{t,s}} \le \phi,$$
$$H_t^s\big(H_s^t(\phi)\big) = \phi^{c_+^{t,s} c_+^{t,s}} \ge \phi,$$

(iii) For any $t,\ s \in [0,1]$ it holds
$$H_t^s \circ H_s^t \circ H_t^s = H_t^s.$$

Proof The order preserving property is a straightforward consequence of the definition. To prove property (ii) observe that

$$H_s^t\big(H_t^s(\phi)\big)(x) = \sup_y \inf_{x'} \Big(\phi(x') + c^{t,s}(x', y) - c^{t,s}(x, y)\Big),$$

which gives the equality $H_s^t\big(H_t^s(\phi)\big) = \phi^{c_-^{t,s} c_-^{t,s}}$: in particular, choosing $x' = x$ we get the claim (the proof of the other equation is similar). For the last property assume $t < s$ (the other case is similar) and observe that by (i) we have

$$\underbrace{H_t^s \circ H_s^t}_{\ge Id} \circ H_t^s \ge H_t^s$$

and
$$H_t^s \circ \underbrace{H_s^t \circ H_t^s}_{\le Id} \le H_t^s.$$

□

The fact that Kantorovich potentials evolve according to the Hopf-Lax formula is expressed in the following theorem. We remark that in the statement below one must deal at the same time with c-concave and c-convex potentials.

Theorem 2.18 (Interpolation of potentials) *Let (X, d) be a Polish geodesic space, $(\mu_t) \subset \mathscr{P}_2(X)$ a constant speed geodesic in $(\mathscr{P}_2(X), W_2)$ and φ a $c = c^{0,1}$-*convex *Kantorovich potential for the couple (μ_0, μ_1). Then the function $\varphi_s := H_0^s(\varphi)$ is a $c^{t,s}$-*concave *Kantorovich potential for the couple (μ_s, μ_t), for any $t < s$.*

*Similarly, if ϕ is a c-*concave *Kantorovich potential for (μ_1, μ_0), then $H_1^t(\phi)$ is a $c^{t,s}$-*convex *Kantorovich potential for (μ_t, μ_s) for any $t < s$.*

Observe that that for $t = 0,\ s = 1$ the theorem reduces to the fact that $H_0^1(\varphi) = (-\varphi)^{c_+}$ is a c-concave Kantorovich potential for $\mu_1,\ \mu_0$, a fact that was already clear by the symmetry of the dual problem discussed in Section 1.3.

Proof
We will prove only the first part of the statement, as the second is analogous.
Step 1. We prove that $H_0^s(\psi)$ is a $c^{t,s}$-concave function for any $t < s$ and any $\psi : X \to \mathbb{R} \cup \{+\infty\}$. This is a consequence of the equality

$$c^{0,s}(x, y) = \inf_{z \in X} c^{0,t}(z, y) + c^{t,s}(x, z),$$

from which it follows

$$H_0^s(\psi)(x) = \inf_{y\in X} c^{0,s}(x,y) + \psi(y) = \inf_{z\in X} c^{t,s}(x,z) + \left(\inf_{y\in X} c^{0,t}(z,y) + \psi(y)\right).$$

Step 2. Let $\boldsymbol{\mu} \in \mathscr{P}(\mathrm{Geod}(X))$ be a measure associated to the geodesic (μ_t) via equation (2.7). We claim that for every $\gamma \in \mathrm{supp}(\boldsymbol{\mu})$ and $s \in (0,1]$ it holds

$$\varphi_s(\gamma_s) = \varphi(\gamma_0) + c^{0,s}(\gamma_0,\gamma_s). \tag{2.13}$$

Indeed the inequality $\leq$ comes directly from the definition by taking $x = \gamma_0$. To prove the opposite one, observe that since $(\mathrm{e}_0,\mathrm{e}_1)_\#\boldsymbol{\mu} \in \mathit{Opt}(\mu_0,\mu_1)$ and φ is a c-convex Kantorovich potential for μ_0,μ_1, we have from Theorem 1.13 that

$$\varphi^{c_-}(\gamma_1) = -c^{0,1}(\gamma_0,\gamma_1) - \varphi(\gamma_0),$$

thus

$$\begin{aligned}\varphi(x) &= \sup_{y\in X} -c^{0,1}(x,y) - \varphi^{c_-}(y) \geq -c^{0,1}(x,\gamma_1) - \varphi^{c_-}(\gamma_1)\\ &= -c^{0,1}(x,\gamma_1) + c^{0,1}(\gamma_0,\gamma_1) + \varphi(\gamma_0).\end{aligned}$$

Plugging this inequality in the definition of φ_s we get

$$\begin{aligned}\varphi_s(\gamma_s) &= \inf_{x\in X} c^{0,s}(x,\gamma_s) + \varphi(x)\\ &\geq \inf_{x\in X} c^{0,s}(x,\gamma_s) - c^{0,1}(x,\gamma_1) + c^{0,1}(\gamma_0,\gamma_1) + \varphi(\gamma_0)\\ &\geq -c^{s,1}(\gamma_s,\gamma_1) + c^{0,1}(\gamma_0,\gamma_1) - \varphi(\gamma_0) = c^{0,s}(\gamma_0,\gamma_s) + \varphi(\gamma_0).\end{aligned}$$

Step 3. We know that an optimal transport plan from μ_t to μ_s is given by $(\mathrm{e}_t,\mathrm{e}_s)_\#\boldsymbol{\mu}$, thus to conclude the proof we need to show that

$$\varphi_s(\gamma_s) + (\varphi_s)^{c_+^{t,s}}(\gamma_t) = c^{t,s}(\gamma_t,\gamma_s), \qquad \forall\gamma \in \mathrm{supp}(\boldsymbol{\mu}),$$

where $(\varphi_s)^{c_+^{t,s}}$ is the $c^{t,s}$-conjugate of the $c^{t,s}$-concave function φ_s. The inequality $\leq$ follows from the definition of $c^{t,s}$-conjugate. To prove opposite inequality start observing that

$$\begin{aligned}\varphi_s(y) &= \inf_{x\in X} c^{0,s}(x,y) + \varphi(y) \leq c^{0,s}(\gamma_0,y) + \varphi(\gamma_0)\\ &\leq c^{0,t}(\gamma_0,\gamma_t) + c^{t,s}(\gamma_t,y) + \varphi(\gamma_0),\end{aligned}$$

and conclude by

$$\begin{aligned}\varphi_s^{c_+^{t,s}}(\gamma_t) &= \inf_{y\in X} c^{t,s}(\gamma_t,y) - \varphi_s(y) \geq -c^{0,t}(\gamma_0,\gamma_t) - \varphi(\gamma_0)\\ &= -c^{0,s}(\gamma_0,\gamma_s) + c^{t,s}(\gamma_t,\gamma_s) - \varphi(\gamma_0)\\ &\overset{(2.13)}{=} c^{t,s}(\gamma_t,\gamma_s) - \varphi_s(\gamma_s).\end{aligned}$$

□

We conclude the section studying some curvature properties of $(\mathscr{P}_2(X), W_2)$. We will focus on spaces *positively/non positively curved* in the sense of Alexandrov, which are the non smooth analogous of Riemannian manifolds having sectional curvature bounded from below/above by 0.

Definition 2.19 (PC and NPC spaces) *A geodesic space (X, d) is said to be positively curved (PC) in the sense of Alexandrov if for every constant speed geodesic $\gamma : [0, 1] \to X$ and every $z \in X$ the following concavity inequality holds:*

$$d^2\big(\gamma_t, z\big) \geq (1-t)d^2\big(\gamma_0, z\big) + td^2\big(\gamma_1, z\big) - t(1-t)d^2\big(\gamma_0, \gamma_1\big). \tag{2.14}$$

Similarly, X is said to be non positively curved (NPC) in the sense of Alexandrov if the converse inequality always holds.

Observe that in an Hilbert space equality holds in (2.14).

The result here is that $(\mathscr{P}_2(X), W_2)$ is PC if (X, d) is, while in general it is not NPC if X is.

Theorem 2.20 ($(\mathscr{P}_2(X), W_2)$ is PC if (X, d) is) *Assume that (X, d) is positively curved. Then $(\mathscr{P}_2(X), W_2)$ is positively curved as well.*

Proof Let (μ_t) be a constant speed geodesic in $\mathscr{P}_2(X)$ and $\nu \in \mathscr{P}_2(X)$. Let $\boldsymbol{\mu} \in \mathscr{P}_2(\mathrm{Geod}(X))$ be a measure such that

$$\mu_t = (\mathrm{e}_t)_\# \boldsymbol{\mu}, \qquad \forall t \in [0, 1],$$

as in Theorem 2.10. Fix $t_0 \in [0, 1]$ and choose $\boldsymbol{\gamma} \in \mathcal{O}pt(\mu_{t_0}, \nu)$. Using a gluing argument (we omit the details) it is possible to show the existence a measure $\boldsymbol{\alpha} \in \mathscr{P}(\mathrm{Geod}(X) \times X)$ such that

$$\begin{aligned} \pi^{\mathrm{Geod}(X)}_\# \boldsymbol{\alpha} &= \boldsymbol{\mu}, \\ \big(\mathrm{e}_{t_0}, \pi^X\big)_\# \boldsymbol{\alpha} &= \boldsymbol{\gamma}, \end{aligned} \tag{2.15}$$

where $\pi^{\mathrm{Geod}(X)}(\gamma, x) := \gamma \in \mathrm{Geod}(X)$, $\pi^X(\gamma, x) := x \in X$ and $\mathrm{e}_{t_0}(\gamma, x) := \gamma_{t_0} \in X$. Then $\boldsymbol{\alpha}$ satisfies also

$$\begin{aligned} \big(\mathrm{e}_0, \pi^X\big)_\# \boldsymbol{\alpha} &\in \mathcal{A}dm(\mu_0, \nu) \\ \big(\mathrm{e}_1, \pi^X\big)_\# \boldsymbol{\alpha} &\in \mathcal{A}dm(\mu_1, \nu), \end{aligned} \tag{2.16}$$

and therefore it holds

$$\begin{aligned} W_2^2(\mu_{t_0}, \nu) &= \int d^2(\mathrm{e}_{t_0}(\gamma), x) d\boldsymbol{\alpha}(\gamma, x) \\ &\overset{(2.14)}{\geq} \int (1-t_0)d^2\big(\gamma_0, z\big) + t_0 d^2\big(\gamma_1, z\big) - t_0(1-t_0)d^2\big(\gamma_0, \gamma_1\big) d\boldsymbol{\alpha}(\gamma, x) \\ &\overset{(2.15)}{=} (1-t_0)\int d^2\big(\gamma_0, z\big) d\boldsymbol{\alpha}(\gamma, x) + t_0 \int d^2\big(\gamma_1, z\big) d\boldsymbol{\alpha}(\gamma, x) \\ &\qquad - t_0(1-t_0)\int d^2\big(\gamma_0, \gamma_1\big) d\boldsymbol{\mu}(\gamma) \\ &\overset{(2.16)}{\geq} (1-t_0)W_2^2(\mu_0, \nu) + t_0 W_2^2(\mu_1, \nu) - t_0(1-t_0)W_2^2(\mu_0, \mu_1), \end{aligned}$$

and by the arbitrariness of t_0 we conclude. $\square$

Example 2.21 (**$(\mathscr{P}_2(X), W_2)$ may be not NPC if (X, d) is**) Let $X = \mathbb{R}^2$ with the Euclidean distance. We will prove that $(\mathscr{P}_2(\mathbb{R}^2), W_2)$ is not NPC. Define

$$\mu_0 := \frac{1}{2}(\delta_{(1,1)} + \delta_{(5,3)}), \quad \mu_1 := \frac{1}{2}(\delta_{(-1,1)} + \delta_{(-5,3)}), \quad \nu := \frac{1}{2}(\delta_{(0,0)} + \delta_{(0,-4)}),$$

then explicit computations show that $W_2^2(\mu_0, \mu_1) = 40$ and $W_2^2(\mu_0, \nu) = 30 = W_2^2(\mu_1, \nu)$. The unique constant speed geodesic (μ_t) from μ_0 to μ_1 is given by

$$\mu_t = \frac{1}{2}\big(\delta_{(1-6t,1+2t)} + \delta_{(5-6t,3-2t)}\big),$$

and simple computations show that

$$40 = W_2^2(\mu_{1/2}, \nu) > \frac{30}{2} + \frac{30}{2} - \frac{40}{4}.$$

■

2.3 X Riemannian manifold

In this section X will always be a compact, smooth Riemannian manifold M without boundary, endowed with the Riemannian distance d.

We study two aspects: the first one is the analysis of some important consequences of Theorem 2.18 about the structure of geodesics in $\mathscr{P}_2(M)$, the second one is the introduction of the so called *weak Riemannian structure* of $(\mathscr{P}_2(M), W_2)$.

Notice that since M is compact, $\mathscr{P}_2(M) = \mathscr{P}(M)$. Yet, we stick to the notation $\mathscr{P}_2(M)$ because all the statements we make in this section are true also for non compact manifolds (although, for simplicity, we prove them only in the compact case).

2.3.1 Regularity of interpolated potentials and consequences

We start observing how Theorem 2.10 specializes to the case of Riemannian manifolds:

Corollary 2.22 (Geodesics in $(\mathscr{P}_2(M), W_2)$) *Let $(\mu_t) \subset \mathscr{P}_2(M)$. Then the following two things are equivalent:*

i) (μ_t) is a geodesic in $(\mathscr{P}_2(M), W_2)$,

ii) there exists a plan $\boldsymbol{\gamma} \in \mathscr{P}(TM)$ (TM being the tangent bundle of M) such that

$$\begin{aligned} \int |\mathrm{v}|^2 d\boldsymbol{\gamma}(x, \mathrm{v}) &= W_2^2(\mu_0, \mu_1), \\ \big(\mathrm{Exp}(t)\big)_\# \boldsymbol{\gamma} &= \mu_t, \end{aligned} \tag{2.17}$$

$\mathrm{Exp}(t) : TM \to M$ *being defined by* $(x, \mathrm{v}) \mapsto \exp_x(t\mathrm{v})$.

Also, for any $\mu, \nu \in \mathscr{P}_2(M)$ such that μ is a regular measure (Definition 1.32), the geodesic connecting μ to ν is unique.

Notice that we cannot substitute the first equation in (2.17) with $(\pi^M, \exp)_\# \boldsymbol{\gamma} \in \mathit{Opt}(\mu_0, \mu_1)$, because this latter condition is strictly weaker (it may be that the curve $t \mapsto \exp_x(t\mathrm{v})$ is not a globally minimizing geodesic from x to $\exp_x(\mathrm{v})$ for some $(x, \mathrm{v}) \in \operatorname{supp} \boldsymbol{\gamma}$).

Proof The implication $(i) \Rightarrow (ii)$ follows directly from Theorem 2.10 by taking into account the fact that $t \mapsto \gamma_t$ is a constant speed geodesic on M implies that for some $(x, \mathrm{v} \in TM)$ it holds $\gamma_t = \exp_x(t\mathrm{v})$ and in this case $d(\gamma_0, \gamma_1) = |\mathrm{v}|$.

For the converse implication, just observe that from the second equation in (2.17) we have

$$W_2^2(\mu_t, \mu_s) \leq \int d^2\big(\exp_x(t\mathrm{v}), \exp_x(s\mathrm{v})\big) d\boldsymbol{\gamma}(x, \mathrm{v}) \leq (t-s)^2 \int |\mathrm{v}|^2 d\boldsymbol{\gamma}(x, \mathrm{v}) = (t-s)^2 W_2^2(\mu_0, \mu_1),$$

having used the first equation in (2.17) in the last step.

To prove the last claim just recall that by Remark 1.35 we know that for μ-a.e. x there exists a unique geodesic connecting x to $T(x)$, T being the optimal transport map. Hence the conclusion follows from (ii) of Theorem 2.10. □

Now we discuss the regularity properties of Kantorovich potentials which follows from Theorem 2.18.

Corollary 2.23 (Regularity properties of the interpolated potentials) *Let ψ be a $c-$convex potential for (μ_0, μ_1) and let $\varphi := H_0^1(\psi)$. Define $\psi_t := H_0^t(\psi)$, $\varphi_t := H_1^t(\varphi)$ and choose a geodesic (μ_t) from μ_0 to μ_1. Then for every $t \in (0, 1)$ it holds:*

i) $\psi_t \geq \varphi_t$ *and both the functions are real valued,*

ii) $\psi_t = \varphi_t$ *on* $\mathrm{supp}(\mu_t)$,

iii) ψ_t *and* φ_t *are differentiable in the support of* μ_t *and on this set their gradients coincide.*

Proof For (i) we have

$$\varphi_t = H_1^t(\varphi) = (H_1^t \circ H_0^1)(\psi) = (\underbrace{H_1^t \circ H_t^1}_{\leq Id} \circ H_0^t)\psi \leq H_0^t(\psi) = \psi_t.$$

Now observe that by definition, $\psi_t(x) < +\infty$ and $\varphi_t(x) > -\infty$ for every $x \in M$, thus it holds

$$+\infty > \psi_t(x) \geq \varphi_t(x) > -\infty, \qquad \forall x \in M.$$

To prove (ii), let $\boldsymbol{\mu}$ be the unique plan associated to the geodesic (μ_t) via (2.7) (recall Proposition 2.16 for uniqueness) and pick $\gamma \in \mathrm{supp}(\boldsymbol{\mu})$. Recall that it holds

$$\begin{aligned}\psi_t(\gamma_t) &= c^{0,t}(\gamma_0, \gamma_t) + \psi(\gamma_0),\\ \varphi_t(\gamma_t) &= c^{t,1}(\gamma_t, \gamma_1) + \varphi(\gamma_1).\end{aligned}$$

Thus from $\varphi(\gamma_1) = c^{0,1}(\gamma_0, \gamma_1) + \psi(\gamma_0)$ we get that $\psi_t(\gamma_t) = \varphi_t(\gamma_t)$. Since $\mu_t = (\mathrm{e}_t)_\#\boldsymbol{\mu}$, the compactness of M gives $\mathrm{supp}(\mu_t) = \{\gamma_t\}_{\gamma \in \mathrm{supp}(\boldsymbol{\mu})}$, so that (ii) follows.

Now we turn to (iii). With the same choice of $t \mapsto \gamma_t$ as above, recall that it holds

$$\begin{aligned}\psi_t(\gamma_t) &= c^{0,t}(\gamma_0, \gamma_t) + \psi(\gamma_0)\\ \psi_t(x) &\leq c^{0,t}(\gamma_0, x) + \psi(\gamma_0), \qquad \forall x \in M,\end{aligned}$$

and that the function $x \mapsto c^{0,t}(\gamma_0, x) + \psi(\gamma_0)$ is superdifferentiable at $x = \gamma_t$. Thus the function $x \mapsto \psi_t$ is superdifferentiable at $x = \gamma_t$. Similarly, φ_t is subdifferentiable at γ_t. Choose $v_1 \in \partial^+\psi_t(\gamma_t)$, $v_2 \in \partial^-\varphi_t(\gamma_t)$ and observe that

$$\psi_t(\gamma_t) + \big\langle v_1, \exp_{\gamma_t}^{-1}(x)\big\rangle + o(D(x, \gamma_t)) \geq \psi_t(x) \geq \varphi_t(x) \geq \varphi_t(\gamma_t) + \big\langle v_2, \exp_{\gamma_t}^{-1}(x)\big\rangle + o(D(x, \gamma_t)),$$

which gives $v_1 = v_2$ and the thesis. □

Corollary 2.24 (The intermediate transport maps are locally Lipschitz) *Let $(\mu_t) \subset \mathscr{P}_2(M)$ a constant speed geodesic in $(\mathscr{P}_2(M), W_2)$. Then for every $t \in (0,1)$ and $s \in [0,1]$ there exists only one optimal transport plan from μ_t to μ_s, this transport plan is induced by a map, and this map is locally Lipschitz.*

Note: clearly in a compact setting being locally Lipschitz means being Lipschitz. We wrote 'locally' because this is the regularity of transport maps in the non compact situation.

Proof Fix $t \in (0,1)$ and, without loss of generality, let $s = 1$. The fact that the optimal plan from is unique and induced by a map is known by Proposition 2.16. Now let v be the vector field defined on $\mathrm{supp}(\mu_t)$ by $v(x) = \nabla\varphi_t = \nabla\psi_t$ (we are using part (iii) of the above corollary, with the same notation). The fact that ψ_t is a $c^{0,t}$-concave potential for the couple μ_t, μ_0 tells that the optimal transport map T satisfies $T(x) \in \partial^{c_+^{0,t}}\phi_t(x)$ for μ_t-a.e. x. Using Theorem 1.33, the fact that ψ_t is differentiable in $\mathrm{supp}(\mu_t)$ and taking into account the scaling properties of the cost, we get that T may be written as $T(x) = \exp_x -v(x)$. Since the exponential map is C^∞, the fact that T is Lipschitz will follow if we show that the vector field v on $\mathrm{supp}(\mu_t)$ is, when read in charts, Lipschitz.

Thus, passing to local coordinates and recalling that $d^2(\cdot, y)$ is uniformly semiconcave, the situation is the following. We have a semiconcave function $f : \mathbb{R}^d \to \mathbb{R}$ and a semiconvex function $g : \mathbb{R}^d \to \mathbb{R}$ such that $f \geq g$ on $\mathbb{R}^d$, $f = g$ on a certain closed set K and we have to prove that the vector field $u : K \to \mathbb{R}^d$ defined by $u(x) = \nabla f(x) = \nabla g(x)$ is Lipschitz. Up to rescaling we may assume that f and g are such that $f - |\cdot|^2$ is concave and $g + |\cdot|^2$ is convex. Then for every $x \in K$ and $y \in \mathbb{R}^d$ we have

$$\langle u(x), y - x\rangle - |x-y|^2 \leq g(y) - g(x) \leq f(y) - f(x) \leq \langle u(x), y - x\rangle + |y - x|^2,$$

and thus for every $x \in K$, $y \in \mathbb{R}^d$ it holds

$$|f(y) - f(x) - \langle u(x), y - x\rangle\,| \leq |x - y|^2.$$

Picking $x_1, x_2 \in K$ and $y \in \mathbb{R}^d$ we have

$$\begin{gathered} f(x_2) - f(x_1) - \langle u(x_1), x_2 - x_1\rangle \leq |x_1 - x_2|^2, \\ f(x_2 + y) - f(x_2) - \langle u(x_2), y\rangle \leq |y|^2, \\ -f(x_2 + y) + f(x_1) + \langle u(x_1), x_2 + y - x_1\rangle \leq |x_2 + y - x_1|^2. \end{gathered}$$

Adding up we get

$$\langle u(x_1) - u(x_2), y\rangle \leq |x_1 - x_2|^2 + |y|^2 + |x_2 + y - x_1|^2 \leq 3(|x_1 - x_2|^2 + |y|^2).$$

Eventually, choosing $y = (u(x_1) - u(x_2))/6$ we obtain

$$|u(x_1) - u(x_2)|^2 \leq 36|x_1 - x_2|^2.$$

□

It is worth stressing the fact that the regularity property ensured by the previous corollary holds without any assumption on the measures μ_0, μ_1.

Remark 2.25 (A (much) simpler proof in the Euclidean case) The fact that intermediate transport maps are Lipschitz can be proved, in the Euclidean case, via the theory of monotone operators. Indeed if $G : \mathbb{R}^d \to \mathbb{R}^d$ is a - possibly multivalued - monotone map (i.e. satisfies $\langle y_1 - y_2, x_1 - x_2\rangle \geq 0$ for every $x_1, x_2 \in \mathbb{R}^d$, $y_i \in G(x_i)$, $i = 1, 2$), then the operator

$((1-t)Id+tG)^{-1}$ is single valued, Lipschitz, with Lipschitz constant bounded above by $1/(1-t)$. To prove this, pick $x_1,\ x_2 \in \mathbb{R}^d$, $y_1 \in G(x_1)$, $y_2 \in G(x_2)$ and observe that

$$
\begin{aligned}
&|(1-t)x_1 + ty_1 - (1-t)x_2 + ty_2|^2 \\
&\qquad = (1-t)^2|x_1-x_2|^2 + t^2|y_1-y_2|^2 + 2t(1-t)\langle x_1-x_2, y_1-y_2\rangle \geq (1-t)^2|x_1-x_2|^2,
\end{aligned}
$$

which is our claim.

Now pick $\mu_0, \mu_1 \in \mathscr{P}_2(\mathbb{R}^d)$, an optimal plan $\boldsymbol{\gamma} \in \mathit{Opt}(\mu_0,\mu_1)$ and consider the geodesic $t \mapsto \mu_t := ((1-t)\pi^1 + t\pi^2)_\# \boldsymbol{\gamma}$ (recall Remark 2.13). From Theorem 1.26 we know that there exists a convex function φ such that $\mathrm{supp}(\boldsymbol{\gamma}) \subset \partial^-\varphi$. Also, we know that the unique optimal plan from μ_0 to μ_t is given by the formula

$$\big(\pi^1, (1-t)\pi^1 + t\pi^2\big)_\# \boldsymbol{\gamma},$$

which is therefore supported in the graph of $(1-t)Id + t\partial^-\varphi$. Since the subdifferential of a convex function is a monotone operator, the thesis follows from the previous claim.

Considering the case in which μ_1 is a delta and μ_0 is not, we can easily see that the bound $(1-t)^{-1}$ on the Lipschitz constant of the optimal transport map from μ_t to μ_0 is sharp. ■

An important consequence of Corollary 2.24 is the following proposition:

Proposition 2.26 (Geodesic convexity of the set of absolutely continuous measures) *Let M be a Riemannian manifold, $(\mu_t) \subset \mathscr{P}_2(M)$ a geodesic and assume that μ_0 is absolutely continuous w.r.t. the volume measure (resp. gives 0 mass to Lipschitz hypersurfaces of codimension 1). Then μ_t is absolutely continuous w.r.t. the volume measure (resp. gives 0 mass to Lipschitz hypersurfaces of codimension 1) for every $t < 1$. In particular, the set of absolutely continuous measures is geodesically convex (and the same for measures giving 0 mass to Lipschitz hypersurfaces of codimension 1).*

Proof Assume that μ_0 is absolutely continuous, let $A \subset M$ be of 0 volume measure, $t \in (0,1)$ and let T_t be the optimal transport map from μ_t to μ_0. Then for every Borel set $A \subset M$ it holds $T_t^{-1}(T_t(A)) \supset A$ and thus

$$\mu_t(A) \leq \mu_t(T_t^{-1}(T_t(A))) = \mu_0(T_t(A)).$$

The claims follow from the fact that T_t is locally Lipschitz. □

Remark 2.27 (The set of regular measures is *not* geodesically convex) It is natural to ask whether the same conclusion of the previous proposition holds for the set of regular measures (Definitions 1.25 and 1.32). The answer is *not*: there are examples of regular measures μ_0, μ_1 in $\mathscr{P}_2(\mathbb{R}^2)$ such that the middle point of the geodesic connecting them is not regular. ■

2.3.2 The weak Riemannian structure of $(\mathscr{P}_2(M), W_2)$

In order to introduce the weak differentiable structure of $(\mathscr{P}_2(X), W_2)$, we start with some heuristic considerations. Let $X = \mathbb{R}^d$ and (μ_t) be a constant speed geodesic on $\mathscr{P}_2(\mathbb{R}^d)$ induced by some optimal map T, i.e.:

$$\mu_t = \big((1-t)Id + tT\big)_\# \mu_0.$$

Then a simple calculation shows that (μ_t) satisfies the continuity equation

$$\frac{d}{dt}\mu_t + \nabla\cdot(v_t\mu_t) = 0,$$

with $v_t := (T - Id) \circ ((1-t)Id + tT)^{-1}$ for every t, in the sense of distributions. Indeed for $\phi \in C_c^\infty(\mathbb{R}^d)$ it holds

$$\frac{d}{dt}\int \phi d\mu_t = \frac{d}{dt}\int \phi\big((1-t)Id+tT\big)d\mu_0 = \int \langle \nabla\phi\big((1-t)Id+tT\big), T-Id\rangle\, d\mu_0 = \int \langle \nabla\phi, v_t\rangle d\mu_t.$$

Now, the continuity equation describes the link between the motion of the continuum μ_t and the instantaneous velocity $v_t : \mathbb{R}^d \to \mathbb{R}^d$ of every "atom" of μ_t. It is therefore natural to think at the vector field v_t as the infinitesimal variation of the continuum μ_t.

From this perspective, one might expect that the set of "smooth" curves on $\mathscr{P}_2(\mathbb{R}^d)$ (and more generally on $\mathscr{P}_2(M)$) is somehow linked to the set of solutions of the continuity equation. This is actually the case, as we are going to discuss now.

In order to state the rigorous result, we need to recall the definition of *absolutely continuous curve* on a metric space.

Definition 2.28 (Absolutely continuous curve) *Let $(Y, \tilde{d})$ be a metric space and let $[0,1] \ni t \mapsto y_t \in Y$ be a curve. Then (y_t) is said absolutely continuous if there exists a function $f \in L^1(0,1)$ such that*

$$\tilde{d}(y_t, y_s) \le \int_t^s f(r)dr, \qquad \forall t < s \in [0,1]. \tag{2.18}$$

We recall that if (y_t) is absolutely continuous, then for a.e. t the *metric derivative* $|\dot{y}_t|$ exists, given by

$$|\dot{y}_t| := \lim_{h\to 0} \frac{\tilde{d}(y_{t+h}, y_t)}{|h|}, \tag{2.19}$$

and that $|\dot{y}_t| \in L^1(0,1)$ and is the smallest L^1 function (up to negligible sets) for which inequality (2.18) is satisfied (see e.g. Theorem 1.1.2 of [6] for the simple proof).

The link between absolutely continuous curves in $\mathscr{P}_2(M)$ and the continuity equation is given by the following theorem:

Theorem 2.29 (Characterization of absolutely continuous curves in $(\mathscr{P}_2(M), W_2)$) *Let M be a smooth complete Riemannian manifold without boundary. Then the following holds.*
*(**A**) For every absolutely continuous curve $(\mu_t) \subset \mathscr{P}_2(M)$ there exists a Borel family of vector fields v_t on M such that $\|v_t\|_{L^2(\mu_t)} \le |\dot{\mu}_t|$ for a.e. t and the continuity equation*

$$\frac{d}{dt}\mu_t + \nabla\cdot(v_t\mu_t) = 0, \tag{2.20}$$

holds in the sense of distributions.
*(**B**) If (μ_t, v_t) satisfies the continuity equation (2.20) in the sense of distributions and $\int_0^1 \|v_t\|_{L^2(\mu_t)}dt < \infty$, then up to redefining $t \mapsto \mu_t$ on a negligible set of times, (μ_t) is an absolutely continuous curve on $\mathscr{P}_2(M)$ and $|\dot{\mu}_t| \le \|v_t\|_{L^2(\mu_t)}$ for a.e. $t \in [0,1]$.*

Note that we are not assuming any kind of regularity on the μ_t's.

We postpone the (sketch of the) proof of this theorem to the end of the section, for the moment we analyze its consequences in terms of the geometry of $\mathscr{P}_2(M)$.

The first important consequence is that the Wasserstein distance, which was defined via the 'static' optimal transport problem, can be recovered via the following 'dynamic' Riemannian-like formula:

Proposition 2.30 (Benamou-Brenier formula) *Let $\mu^0,\, \mu^1 \in \mathscr{P}_2(M)$. Then it holds*

$$W_2(\mu^0, \mu^1) = \inf \left\{ \int_0^1 \|v_t\|_{\mu_t} dt \right\}, \tag{2.21}$$

where the infimum is taken among all weakly continuous distributional solutions of the continuity equation (μ_t, v_t) such that $\mu_0 = \mu^0$ and $\mu_1 = \mu^1$.

Proof We start with inequality $\leq$. Let (μ_t, v_t) be a solution of the continuity equation. Then if $\int_0^1 \|v_t\|_{L^2(\mu_t)} = +\infty$ there is nothing to prove. Otherwise we may apply part **B** of Theorem 2.29 to get that (μ_t) is an absolutely continuous curve on $\mathscr{P}_2(M)$. The conclusion follows from

$$W_2(\mu^0, \mu^1) \leq \int_0^1 |\dot{\mu}_t| dt \leq \int_0^1 \|v_t\|_{L^2(\mu_t)} dt,$$

where in the last step we used part (**B**) of Theorem 2.29 again.

To prove the converse inequality it is enough to consider a constant speed geodesic (μ_t) connecting μ^0 to μ^1 and apply part (**A**) of Theorem 2.29 to get the existence of vector fields v_t such that the continuity equation is satisfied and $\|v_t\|_{L^2(\mu_t)} \leq |\dot{\mu}_t| = W_2(\mu^0, \mu^1)$ for a.e. $t \in [0, 1]$. Then we have

$$W_2(\mu^0, \mu^1) \geq \int_0^1 \|v_t\|_{L^2(\mu_t)} dt,$$

as desired. □

This proposition strongly suggests that the scalar product in $L^2(\mu)$ should be considered as the metric tensor on $\mathscr{P}_2(M)$ at μ. Now observe that given an absolutely continuous curve $(\mu_t) \subset \mathscr{P}_2(M)$ in general there is no unique choice of vector field (v_t) such that the continuity equation (2.20) is satisfied. Indeed, if (2.20) holds and w_t is a Borel family of vector fields such that $\nabla \cdot (w_t \mu_t) = 0$ for a.e. t, then the continuity equation is satisfied also with the vector fields $(v_t + w_t)$. It is then natural to ask whether there is some natural selection principle to associate uniquely a family of vector fields (v_t) to a given absolutely continuous curve. There are two possible approaches:
Algebraic approach. The fact that for distributional solutions of the continuity equation the vector field v_t acts only on gradients of smooth functions suggests that the v_t's should be taken in the set of gradients as well, or, more rigorously, v_t should belong to

$$\overline{\left\{ \nabla \varphi \ : \ \varphi \in C_c^\infty(M) \right\}}^{L^2(\mu_t)} \tag{2.22}$$

for a.e. $t \in [0, 1]$.
Variational approach. The fact that the continuity equation is linear in v_t and the L^2 norm is strictly convex, implies that there exists a unique, up to negligible sets in time, family of vector fields $v_t \in L^2(\mu_t)$, $t \in [0, 1]$, with minimal norm for a.e. t, among the vector fields compatible with the curve (μ_t) via the continuity equation. In other words, for any other vector field $(\tilde{v}_t)$ compatible with the curve (μ_t) in the sense that (2.20) is satisfied, it holds $\|\tilde{v}_t\|_{L^2(\mu_t)} \geq \|v_t\|_{L^2(\mu_t)}$ for a.e. t. It is immediate to verify that v_t is of minimal norm if and only if it belongs to the set

$$\left\{ v \in L^2(\mu_t) \ : \ \int \langle v, w \rangle \, d\mu_t = 0, \ \forall w \in L^2(\mu_t) \ s.t. \ \nabla \cdot (w \mu_t) = 0 \right\}. \tag{2.23}$$

The important point here is that the sets defined by (2.22) and (2.23) are the same, as it is easy to check. Therefore it is natural to give the following

Definition 2.31 (The tangent space) *Let $\mu \in \mathscr{P}_2(M)$. Then the tangent space $\mathrm{Tan}_\mu(\mathscr{P}_2(M))$ at $\mathscr{P}_2(M)$ in μ is defined as*

$$\begin{aligned}\mathrm{Tan}_\mu(\mathscr{P}_2(M)) :=& \overline{\Big\{\nabla\varphi \ : \ \varphi \in C_c^\infty(M)\Big\}}^{L^2(\mu)}\\ =& \Big\{v \in L^2(\mu) \ : \ \int \langle v, w\rangle \, d\mu = 0, \ \forall w \in L^2(\mu) \ s.t. \ \nabla\cdot(w\mu) = 0\Big\}\end{aligned}$$

Thus we now have a definition of tangent space for every $\mu \in \mathscr{P}_2(M)$ and this tangent space is naturally endowed with a scalar product: the one of $L^2(\mu)$. This fact, Theorem 2.29 and Proposition 2.30 are the bases of the so-called weak Riemannian structure of $(\mathscr{P}_2(M), W_2)$.

We now state, without proof, some other properties of $(\mathscr{P}_2(M), W_2)$ which resemble those of a Riemannian manifold. For simplicity, we will deal with the case $M = \mathbb{R}^d$ only and we will assume that the measures we are dealing with are regular (Definition 1.25), but analogous statements hold for general manifolds and general measures.

In the next three propositions (μ_t) is an absolutely continuous curve in $\mathscr{P}_2(\mathbb{R}^d)$ such that μ_t is regular for every t. Also (v_t) is the unique, up to a negligible set of times, family of vector fields such that the continuity equation holds and $v_t \in \mathrm{Tan}_{\mu_t}(\mathscr{P}_2(\mathbb{R}^d))$ for a.e. t.

Proposition 2.32 (v_t can be recovered by infinitesimal displacement) *Let (μ_t) and (v_t) as above. Also, let T_t^s be the optimal transport map from μ_t to μ_s (which exists and is unique by Theorem 1.26, due to our assumptions on μ_t). Then for a.e. $t \in [0,1]$ it holds*

$$v_t = \lim_{s\to t} \frac{T_t^s - Id}{s-t},$$

the limit being understood in $L^2(\mu_t)$.

Proposition 2.33 ("Displacement tangency") *Let (μ_t) and (v_t) as above. Then for a.e. $t \in [0,1]$ it holds*

$$\lim_{h\to 0} \frac{W_2\big(\mu_{t+h}, (Id + hv_t)_\#\mu_t\big)}{h} = 0. \tag{2.24}$$

Proposition 2.34 (Derivative of the squared distance) *Let (μ_t) and (v_t) as above and $\nu \in \mathscr{P}_2(\mathbb{R}^d)$. Then for a.e. $t \in [0,1]$ it holds*

$$\frac{d}{dt} W_2^2(\mu_t, \nu) = -2 \int \langle v_t, T_t - Id\rangle \, d\mu_t,$$

where T_t is the unique optimal transport map from μ_t to ν (which exists and is unique by Theorem 1.26, due to our assumptions on μ_t).

We conclude the section with a sketch of the proof of Theorem 2.29.

Sketch of the Proof of Theorem 2.29
Reduction to the Euclidean case Suppose we already know the result for the case $\mathbb{R}^d$ and we want to prove it for a compact and smooth manifold M. Use the Nash embedding theorem to get the existence of a smooth map $i : M \to \mathbb{R}^D$ whose differential provides an isometry of T_xM and its image for any $x \in M$. Now notice that the inequality $|i(x) - i(y)| \leq d(x,y)$ valid for any $x, y \in M$ ensures that $W_2(i_\#\mu, i_\#\nu) \leq W_2(\mu,\nu)$ for any $\mu, \nu \in \mathscr{P}_2(M)$. Hence given an absolutely continuous curve $(\mu_t) \subset \mathscr{P}_2(M)$, the curve $(i_\#\mu_t) \subset \mathscr{P}_2(\mathbb{R}^D)$ is absolutely continuous as well, and there exists a family vector fields v_t such that (2.20) is fulfilled with $i_\#\mu_t$ in place of μ_t

and $\|v_t\|_{L^2(i_\# \mu_t)} \le |\dot{i_\# \mu_t}| \le |\dot\mu_t|$ for a.e. t. Testing the continuity equation with functions constant on $i(M)$ we get that for a.e. t the vector field v_t is tangent to $i(M)$ for $i_\#\mu_t$-a.e. point. Thus the v_t's are the (isometric) image of vector fields on M and part (A) is proved.

Viceversa, let $(\mu_t) \subset \mathscr{P}_2(M)$ be a curve and the v_t's vector fields in M such that $\int_0^1 \|v_t\|_{L^2(\mu_t)} dt < \infty$ and assume that they satisfy the continuity equation. Then the measures $\tilde\mu_t := i_\#\mu_t$ and the vector fields $\tilde v_t := di(v_t)$ satisfy the continuity equation on $\mathbb{R}^D$. Therefore $(\tilde\mu_t)$ is an absolutely continuous curve and it holds $|\dot{\tilde\mu}_t| \le \|\tilde v_t\|_{L^2(\tilde\mu_t)} = \|v_t\|_{L^2(\mu_t)}$ for a.e. t. Notice that i is bilipschitz and therefore (μ_t) is absolutely continuous as well. Hence to conclude it is sufficient to show that $|\dot{\tilde\mu}_t| = |\dot\mu_t|$ a.e. t. To prove this, one can notice that the fact that i is bilipschitz and validity of

$$\lim_{r\to 0} \sup_{\substack{x,y\in M\\ d(x,y)<r}} \frac{d(x,y)}{|i(x)-i(y)|} = 1,$$

give that

$$\lim_{r\to 0} \sup_{\substack{\mu,\nu\in \mathscr{P}_2(M)\\ W_2(\mu,\nu)<r}} \frac{W_2(\mu,\nu)}{W_2(i_\#\mu, i_\#\nu)} = 1.$$

We omit the details.

Part A. Fix $\varphi \in C_c^\infty(\mathbb{R}^d)$ and observe that for every $\boldsymbol{\gamma}_t^s \in \mathit{Opt}(\mu_t,\mu_s)$ it holds

$$\begin{aligned}
\left|\int \varphi d\mu_s - \int \varphi d\mu_t\right| &= \left|\int \varphi(y) d\boldsymbol{\gamma}_t^s(x,y) - \int \varphi(x) d\boldsymbol{\gamma}_t^s(x,y)\right| \\
&= \left|\int \varphi(y) - \varphi(x) d\boldsymbol{\gamma}_t^s(x,y)\right| \\
&= \left|\int \int_0^1 \langle \nabla\varphi(x+\lambda(y-x)), y-x\rangle \, d\lambda d\boldsymbol{\gamma}_t^s(x,y)\right| \\
&= \left|\int \langle \nabla\varphi(x), y-x\rangle \, d\boldsymbol{\gamma}_t^s(x,y)\right| + \mathrm{Rem}(\varphi,t,s) \\
&\le \sqrt{\int |\nabla\varphi(x)|^2 d\boldsymbol{\gamma}_t^s(x,y)} \sqrt{\int |x-y|^2 d\boldsymbol{\gamma}_t^s(x,y)} + \mathrm{Rem}(\varphi,t,s) \\
&= \|\nabla\varphi\|_{L^2(\mu_t)} W_2(\mu_t,\mu_s) + \mathrm{Rem}(\varphi,t,s),
\end{aligned} \tag{2.25}$$

where the remainder term $\mathrm{Rem}(\varphi,t,s)$ can be bounded by by

$$|\mathrm{Rem}(\varphi,t,s)| \le \frac{\mathrm{Lip}(\nabla\varphi)}{2} \int |x-y|^2 d\boldsymbol{\gamma}_t^s(x,y) = \frac{\mathrm{Lip}(\nabla\varphi)}{2} W_2^2(\mu_t,\mu_s).$$

Thus (2.25) implies that the map $t \mapsto \int \varphi d\mu_t$ is absolutely continuous for any $\varphi \in C_c^\infty(\mathbb{R}^d)$.

Now let $D \subset C_c^\infty(\mathbb{R}^d)$ be a countable set such that $\{\nabla\varphi : \varphi \in D\}$ is dense in $\mathrm{Tan}_{\mu_t}(\mathscr{P}_2(\mathbb{R}^d))$ for every $t \in [0,1]$ (the existence of such D follows from the compactness of $\{\mu_t\}_{t\in[0,1]} \subset \mathscr{P}_2(\mathbb{R}^d)$, we omit the details). The above arguments imply that there exists a set $A \subset [0,1]$ of full Lebesgue measure such that $t \mapsto \int \varphi d\mu_t$ is differentiable at $t \in A$ for every $\varphi \in D$; we can also assume that the metric derivative $|\dot\mu_t|$ exists for every $t \in A$. Also, by (2.25) we know that for $t_0 \in A$ the linear functional $L_{t_0} : \{\nabla\varphi : \varphi \in D\} \to \mathbb{R}$ given by

$$\nabla\varphi \mapsto L_{t_0}(\nabla\varphi) := \frac{d}{dt}_{|t=t_0} \int \varphi d\mu_t$$

satisfies

$$|L_{t_0}(\nabla\varphi)| \le \|\nabla\varphi\|_{L^2(\mu_{t_0})}|\dot\mu_{t_0}|,$$

and thus it can be uniquely extended to a linear and bounded functional on $\mathrm{Tan}_{\mu_{t_0}}(\mathscr{P}_2(\mathbb{R}^d))$. By the Riesz representation theorem there exists a vector field $v_{t_0} \in \mathrm{Tan}_{\mu_{t_0}}(\mathscr{P}_2(\mathbb{R}^d))$ such that

$$\frac{d}{dt}|_{t=t_0} \int \varphi d\mu_t = L_{t_0}(\nabla\varphi) = \int \langle \nabla\varphi, v_{t_0}\rangle \, d\mu_{t_0}, \qquad \forall \varphi \in D, \tag{2.26}$$

and whose norm in $L^2(\mu_{t_0})$ is bounded above by the metric derivative $|\dot\mu_t|$ at $t = t_0$. It remains to prove that the continuity equation is satisfied in the sense of distributions. This is a consequence of (2.26), see Theorem 8.3.1 of [6] for the technical details.

Part B. Up to a time reparametrization argument, we can assume that $\|v_t\|_{L^2(\mu_t)} \le L$ for some $L \in \mathbb{R}$ for a.e. t. Fix a Gaussian family of mollifiers ρ^ε and define

$$\mu_t^\varepsilon := \mu_t * \rho^\varepsilon,$$
$$v_t^\varepsilon := \frac{(v_t\mu_t) * \rho^\varepsilon}{\mu_t^\varepsilon}.$$

It is clear that

$$\frac{d}{dt}\mu_t^\varepsilon + \nabla\cdot(v_t^\varepsilon \mu_t^\varepsilon) = 0.$$

Moreover, from Jensen inequality applied to the map $(X, z) \mapsto z|X/z|^2 = |X|^2/z$ ($X = v_t\mu_t$) it follows that

$$\|v_t^\varepsilon\|_{L^2(\mu_t^\varepsilon)} \le \|v_t\|_{L^2(\mu_t)} \le L. \tag{2.27}$$

This bound, together with the smoothness of v_t^ε, implies that there exists a unique locally Lipschitz map $\mathbf{T}^\varepsilon(\cdot,\cdot) : [0,1]\times\mathbb{R}^d \to \mathbb{R}^d$, $t \in [0,1]$ satisfying

$$\begin{cases} \frac{d}{dt}\mathbf{T}^\varepsilon(t,x) &= v_t^\varepsilon\big(\mathbf{T}^\varepsilon(t,x)\big) \qquad \forall x \in \mathbb{R}^d,\ a.e.\ t \in [0,1], \\ \mathbf{T}^\varepsilon(t,x) &= x, \qquad \forall x \in \mathbb{R}^d,\ t \in [0,1]. \end{cases}$$

A simple computation shows that the curve $t \mapsto \tilde\mu_t^\varepsilon := \mathbf{T}^\varepsilon(t,\cdot)_\# \mu_0^\varepsilon$ solves

$$\frac{d}{dt}\tilde\mu_t^\varepsilon + \nabla\cdot(v_t^\varepsilon\tilde\mu_t^\varepsilon) = 0, \tag{2.28}$$

which is the same equation solved by (μ_t^ε). It is possible to show that this fact together with the smoothness of the v_t^ε's and the equality $\mu_0^\varepsilon = \tilde\mu_0^\varepsilon$ gives that $\tilde\mu_t^\varepsilon = \mu_t^\varepsilon$ for every t, ε (see Proposition 8.1.7 and Theorem 8.3.1 of [6] for a proof of this fact).

Conclude observing that

$$\begin{aligned} W_2^2(\mu_t^\varepsilon, \mu_s^\varepsilon) &\le \int |\mathbf{T}^\varepsilon(t,x) - \mathbf{T}^\varepsilon(s,x)|^2 d\mu_0^\varepsilon(x) = \int \left| \int_t^s v_r^\varepsilon\big(\mathbf{T}^\varepsilon(r,x)\big)\right|^2 d\mu_0^\varepsilon(x) \\ &\le |t-s| \int\int_t^s \big|v_r^\varepsilon\big(\mathbf{T}^\varepsilon(r,x)\big)\big|^2 dr\, d\mu_0^\varepsilon = |t-s|\int_t^s \big\|v_r^\varepsilon\big(\mathbf{T}^\varepsilon(r,\cdot)\big)\big\|^2_{L^2(\mu_0^\varepsilon)} dr \\ &\le |t-s|\int_t^s \|v_r^\varepsilon\|^2_{L^2(\mu_r^\varepsilon)} dr \overset{(2.27)}{\le} |t-s|^2 L, \end{aligned}$$

and that, by the characterization of convergence (2.4), $W_2(\mu_t^\varepsilon, \mu_t) \to 0$ as $\varepsilon \to 0$ for every $t \in [0,1]$.
$\square$

2.4 Bibliographical notes

To call the distance W_2 the 'Wasserstein distance' is quite not fair: a much more appropriate would be Kantorovich distance. Also, the spelling 'Wasserstein' is questionable, as the original one was 'Vasershtein'. Yet, this terminology is nowadays so common that it would be impossible to change it.

The equivalence (2.4) has been proven by the authors and G. Savaré in [6]. In the same reference Remark 2.8 has been first made. The fact that $(\mathscr{P}_2(X), W_2)$ is complete and separable as soon as (X, d) is belongs to the folklore of the theory, a proof can be found in [6]. Proposition 2.4 was proved by C. Villani in [79], Theorem 7.12.

The terminology *displacement interpolation* was introduced by McCann [63] for probability measures in $\mathbb{R}^d$. Theorem 2.10 appears in this form here for the first time: in [58] the theorem was proved in the compact case, in [80] (Theorem 7.21) this has been extended to locally compact structures and much more general forms of interpolation. The main source of difficulty when dealing with general Polish structure is the potential lack of tightness: the proof presented here is strongly inspired by the work of S. Lisini [54].

Proposition 2.16 and Theorem 2.18 come from [80] (Corollary 7.32 and Theorem 7.36 respectively). Theorem 2.20 and the counterexample 2.21 are taken from [6] (Theorem 7.3.2 and Example 7.3.3 respectively).

The proof of Corollary 2.24 is taken from an argument by A. Fathi [35], the paper being inspired by Bernand-Buffoni [13]. Remark 2.27 is due to N. Juillet [48].

The idea of looking at the transport problem as dynamical problem involving the continuity equation is due to J.D. Benamou and Y. Brenier ([12]), while the fact that $(\mathscr{P}_2(\mathbb{R}^d), W_2)$ can be viewed as a sort of infinite dimensional Riemannian manifold is an intuition by F. Otto [67]. Theorem 2.29 has been proven in [6] (where also Propositions 2.32, 2.33 and 2.34 were proven) in the case $M = \mathbb{R}^d$, the generalization to Riemannian manifolds comes from Nash's embedding theorem.

3 Gradient flows

The aim of this Chapter is twofold: on one hand we give an overview of the theory of Gradient Flows in a metric setting, on the other hand we discuss the important application of the abstract theory to the case of geodesically convex functionals on the space $(\mathscr{P}_2(\mathbb{R}^d), W_2)$.

Let us recall that for a smooth function $F : M \to \mathbb{R}$ on a Riemannian manifold, a gradient flow (x_t) starting from $\overline{x} \in M$ is a differentiable curve solving

$$\begin{cases} x_t' &= -\nabla F(x_t), \\ x_0 &= \overline{x}. \end{cases} \tag{3.1}$$

Observe that there are two necessary ingredients in this definition: the functional F and the metric on M. The role of the functional is clear. The metric is involved to define ∇F: it is used to identify the cotangent vector dF with the tangent vector ∇F.

3.1 Hilbertian theory of gradient flows

In this section we quickly recall the main results of the theory of Gradient flow for λ-convex functionals on Hilbert spaces. This will deserve as guideline for the analysis that we will make later on of the same problem in a purely metric setting.

Let H be Hilbert and $\lambda \in \mathbb{R}$. A λ-convex functional $F : H \to \mathbb{R} \cup \{+\infty\}$ is a functional satisfying:

$$F\big((1-t)x + ty\big) \le (1-t)F(x) + tF(y) - \frac{\lambda}{2}t(1-t)|x-y|^2, \qquad \forall x, y \in H,$$

(this corresponds to $\nabla^2 F \ge \lambda Id$ for functionals on $\mathbb{R}^d$). We denote with $D(F)$ the domain of F, i.e. $D(F) := \{x : F(x) < \infty\}$.

The subdifferential $\partial^- F(x)$ of F at a point $x \in D(F)$ is the set of $v \in H$ such that

$$F(x) + \langle v, y - x\rangle + \frac{\lambda}{2}|x-y|^2 \le F(y), \qquad \forall y \in H.$$

An immediate consequence of the definition is the fact that the subdifferential of F satisfies the *monotonicity inequality*:

$$\langle v - w, x - y\rangle \ge \lambda |x-y|^2 \qquad \forall v \in \partial F(x),\ w \in \partial^- F(y).$$

We will denote by $\nabla F(x)$ the element of minimal norm in $\partial F(x)$, which exists and is unique as soon as $\partial^- F(x) \neq \emptyset$, because $\partial^- F(x)$ is closed and convex.

For convex functions a natural generalization of Definition (3.1) of Gradient Flow is possible: we say that (x_t) is a Gradient Flow for F starting from $\overline{x} \in H$ if it is a locally absolutely continuous curve in $(0, +\infty)$ such that

$$\begin{cases} x'_t \in -\partial^- F(x_t) & \text{for a.e. } t > 0 \\ \lim\limits_{t\downarrow 0} x_t = \overline{x}. \end{cases} \tag{3.2}$$

We now summarize without proof the main existence and uniqueness results in this context.

Theorem 3.1 (Gradient Flows in Hilbert spaces - (Brezis, Pazy)) *If $F : H \to \mathbb{R} \cup \{+\infty\}$ is λ-convex and lower semicontinuous, then the following statements hold.*

(i) **Existence and uniqueness** *for all $\bar{x} \in \overline{D(F)}$* (3.2) *has a unique solution (x_t).*

(ii) **Minimal selection and Regularizing effects** *It holds $\frac{d_+}{dt} x_t = -\nabla F(x_t)$ for every $t > 0$ (that is, the right derivative of x_t always exists and realizes the element of minimal norm in $\partial^- F(x_t)$) and $\frac{d_+}{dt} F \circ x(t) = -|\nabla F(x(t))|^2$ for every $t > 0$. Also*

$$F(x_t) \le \inf_{v \in D(F)} \left\{ F(v) + \frac{1}{2t}|v - \bar{x}|^2 \right\},$$

$$|\nabla F(x_t)|^2 \le \inf_{v \in D(\partial F)} \left\{ |\nabla F(v)|^2 + \frac{1}{t^2}|v - \bar{x}|^2 \right\}.$$

(iii) **Energy Dissipation Equality** *$|x'_t|, |\nabla F|(x_t) \in L^2_{\mathrm{loc}}(0, +\infty)$, $F(x_t) \in AC_{\mathrm{loc}}(0, +\infty)$ and the following Energy Dissipation Equality holds:*

$$F(x_t) - F(x_s) = \frac{1}{2}\int_t^s |\nabla F(x_r)|^2\, dr + \frac{1}{2}\int_t^s |x'_r|^2\, dr \qquad 0 < t \le s < \infty;$$

(iv) **Evolution Variational Inequality and contraction** *(x_t) is the unique solution of the system of differential inequalities*

$$\frac{1}{2}\frac{d}{dt}|\tilde{x}_t - y|^2 + F(x_t) + \frac{\lambda}{2}|\tilde{x}_t - y|^2 \le F(y), \qquad \forall y \in H,\ a.e.\ t,$$

among all locally absolutely continuous curves $(\tilde{x}_t)$ in $(0,\infty)$ converging to $\overline{x}$ as $t \to 0$. Furthermore, if (y_t) is a solution of (3.2) *starting from $\overline{y}$, it holds*

$$|x_t - y_t| \le e^{-\lambda t}|\overline{x} - \overline{y}|.$$

(v) **Asymptotic behavior** *If $\lambda > 0$ then there exists a unique minimum $x_{\min}$ of F and it holds*

$$F(x_t) - F(x_{\min}) \le \big(F(\bar{x}) - F(x_{\min})\big)e^{-2\lambda t}.$$

In particular, the pointwise energy inequality

$$F(x) \ge F(x_{\min}) + \frac{\lambda}{2}|x - x_{\min}|^2, \qquad \forall x \in H$$

gives

$$|x_t - x_{\min}| \le \sqrt{\frac{2(F(\overline{x}) - F(x_{\min}))}{\lambda}}e^{-\lambda t}.$$

3.2 The theory of Gradient Flows in a metric setting

Here we give an overview of the theory of Gradient Flows in a purely metric framework.

3.2.1 The framework

The first thing we need to understand is the meaning of Gradient Flow in a metric setting. Indeed, the system (3.2) makes no sense in metric spaces, thus we need to reformulate it so that it has a metric analogous. There are several ways to do this, below we summarize the most important ones.

For the purpose of the discussion below, we assume that $H = \mathbb{R}^d$ and that $E : H \to \mathbb{R}$ is λ-convex and of class C^1.

Let us start observing that (3.2) may be written as: $t \mapsto x_t$ is locally absolutely continuous in $(0, +\infty)$, converges to $\overline{x}$ as $t \downarrow 0$ and it holds

$$\frac{d}{dt}E\big(x_t\big) \le -\frac{1}{2}|\nabla E|^2\big(x_t\big) - \frac{1}{2}|x_t'|^2, \qquad a.e.\ t \ge 0. \tag{3.3}$$

Indeed, along *any* absolutely continuous curve y_t it holds

$$\begin{aligned}\frac{d}{dt}E\big(y_t\big) &= \langle \nabla E(y_t), y_t' \rangle \\ &\ge -|\nabla E|(y_t)|y_t'| \qquad (= \text{ if and only if } -y_t' \text{ is a positive multiple of } \nabla E(y_t)), \\ &\ge -\frac{1}{2}|\nabla E|^2\big(y_t\big) - \frac{1}{2}|y_t'|^2 \qquad (= \text{ if and only if } |y_t'| = |\nabla E(y_t)|).\end{aligned} \tag{3.4}$$

Thus in particular equation (3.3) may be written in the following integral form

$$E\big(x_s\big) + \frac{1}{2}\int_t^s |x_r'|^2 dr + \frac{1}{2}\int_t^s |\nabla E|^2(x_r) dr \le E(x_t), \qquad a.e.\ t < s \tag{3.5}$$

which we call *Energy Dissipation Inequality* (EDI in the following).

Since the inequality (3.4) shows that $\frac{d}{dt}E\big(y_t\big) < -\frac{1}{2}|\nabla E|^2\big(y_t\big) - \frac{1}{2}|y_t'|^2$ never holds, the system (3.2) may be also written in form of *Energy Dissipation Equality* (EDE in the following) as

$$E\big(x_t\big) + \frac{1}{2}\int_t^s |x_r'|^2 dr + \frac{1}{2}\int_t^s |\nabla E|^2(x_r) dr = E(x_t), \qquad \forall 0 \le t \le s. \tag{3.6}$$

Notice that the convexity of E does not play any role in this formulation.

A completely different way to rewrite (3.2) comes from observing that if x_t solves (3.2) and $y \in H$ is a generic point it holds

$$\frac{1}{2}\frac{d}{dt}|x_t - y|^2 = \langle x_t - y, x'_t \rangle = \langle y - x_t, \nabla E(x_t) \rangle \le E(y) - E(x_t) - \frac{\lambda}{2}|x_t - y|^2,$$

where in the last inequality we used the fact that E is λ-convex. Since the inequality

$$\langle y - x, v \rangle \le E(y) - E(x) - \frac{\lambda}{2}|x - y|^2, \qquad \forall y \in H,$$

characterizes the elements v of the subdifferential of E at x, we have that an absolutely continuous curve x_t solves (3.2) if and only if

$$\frac{1}{2}\frac{d}{dt}|x_t - y|^2 + \frac{1}{2}\lambda|x_t - y|^2 + E(x_t) \le E(y), \qquad a.e.\ t \ge 0, \tag{3.7}$$

holds for every $y \in H$. We will call this system of inequalities the *Evolution Variational Inequality* (EVI).

Thus we got three different characterizations of Gradient Flows in Hilbert spaces: the EDI, the EDE and the EVI. We now want to show that it is possible to formulate these equations also for functionals E defined on a metric space (X, d).

The object $|x'_t|$ appearing in EDI and EDE can be naturally interpreted as the *metric speed* of the absolutely continuous curve x_t as defined in (2.19). The metric analogous of $|\nabla E|(x)$ is the *slope* of E, defined as:

Definition 3.2 (Slope) *Let $E : X \to \mathbb{R} \cup \{+\infty\}$ and $x \in X$ be such that $E(x) < \infty$. Then the slope $|\nabla E|(x)$ of E at x is:*

$$|\nabla E|(x) := \varlimsup_{y \to x} \frac{(E(x) - E(y))^+}{d(x, y)} = \max\left\{ \varlimsup_{y \to x} \frac{E(x) - E(y)}{d(x, y)}\ , 0 \right\}.$$

The three definitions of Gradient Flows in a metric setting that we are going to use are:

Definition 3.3 (Energy Dissipation Inequality definition of GF - EDI) *Let $E : X \to \mathbb{R} \cup \{+\infty\}$ and let $\overline{x} \in X$ be such that $E(\overline{x}) < \infty$. We say that $[0, \infty) \ni t \mapsto x_t \in X$ is a Gradient Flow in the EDI sense starting at $\overline{x}$ provided it is a locally absolutely continuous curve, $x_0 = \overline{x}$ and*

$$\begin{aligned} &E(x_s) + \frac{1}{2}\int_0^s |\dot{x}_r|^2 dr + \frac{1}{2}\int_0^s |\nabla E|^2(x_r) dr \le E(\overline{x}), && \forall s \ge 0, \\ &E(x_s) + \frac{1}{2}\int_t^s |\dot{x}_r|^2 dr + \frac{1}{2}\int_t^s |\nabla E|^2(x_r) dr \le E(x_t), && a.e.\ t > 0,\ \forall s \ge t. \end{aligned} \tag{3.8}$$

Definition 3.4 (Energy Dissipation Equality definition of GF - EDE) *Let $E : X \to \mathbb{R} \cup \{+\infty\}$ and let $\overline{x} \in X$ be such that $E(\overline{x}) < \infty$. We say that $[0, \infty) \ni t \mapsto x_t \in X$ is a Gradient Flow in the EDE sense starting at $\overline{x}$ provided it is a locally absolutely continuous curve, $x_0 = \overline{x}$ and*

$$E(x_s) + \frac{1}{2}\int_t^s |\dot{x}_r|^2 dr + \frac{1}{2}\int_t^s |\nabla E|^2(x_r) dr = E(x_t), \qquad \forall 0 \le t \le s. \tag{3.9}$$

Definition 3.5 (Evolution Variation Inequality definition of GF - EVI) *Let $E : X \to \mathbb{R}\cup\{+\infty\}$, $\overline{x} \in \overline{\{E < \infty\}}$ and $\lambda \in \mathbb{R}$. We say that $(0,\infty) \ni t \mapsto x_t \in X$ is a Gradient Flow in the EVI sense (with respect to λ) starting at $\overline{x}$ provided it is a locally absolutely continuous curve in $(0,\infty)$, $x_t \to \overline{x}$ as $t \to 0$ and*

$$E(x_t) + \frac{1}{2}\frac{d}{dt}d^2(x_t, y) + \frac{\lambda}{2}d^2(x_t, y) \le E(y), \qquad \forall y \in X,\ a.e.\ t > 0.$$

There are two basic and fundamental things that one needs understand when studying the problem of Gradient Flows in a metric setting:

1) Although the formulations EDI, EDE and EVI are equivalent for λ-convex functionals on Hilbert spaces, they are *not* equivalent in a metric setting. Shortly said, it holds

$$EVI \qquad \Rightarrow \qquad EDE \qquad \Rightarrow \qquad EDI$$

and typically none of the converse implication holds (see Examples 3.15 and 3.23 below). Here the second implication is clear, for the proof of the first one see Proposition 3.6 below.

2) Whatever definition of Gradient Flow in a metric setting we use, the main problem is to show existence. The main ingredient in almost all existence proofs is the Minimizing Movements scheme, which we describe after Proposition 3.6.

Proposition 3.6 (EVI implies EDE) *Let $E : X \to \mathbb{R} \cup \{+\infty\}$ be a lower semicontinuous functional, $\overline{x} \in X$ a given point, $\lambda \in \mathbb{R}$ and assume that (x_t) is a Gradient Flow for E starting from $\overline{x}$ in the EVI sense w.r.t. λ. Then equation* (3.9) *holds.*

Proof First we assume that x_t is locally Lipschitz. The claim will be proved if we show that $t \mapsto E(x_t)$ is locally Lipschitz and it holds

$$-\frac{d}{dt}E(x_t) = \frac{1}{2}|\dot{x}_t|^2 + \frac{1}{2}|\nabla E|^2(x_t), \qquad a.e.\ t > 0.$$

Let us start observing that the triangle inequality implies

$$\frac{1}{2}\frac{d}{dt}d^2(x_t, y) \ge -|\dot{x}_t| d(x_t, y), \qquad \forall y \in X,\ a.e.\ t > 0,$$

thus plugging this bound into the EVI we get

$$-|\dot{x}_t| d(x_t, y) + \frac{\lambda}{2}d^2(x_t, y) + E(x_t) \le E(y), \qquad \forall y \in X,\ a.e.\ t > 0,$$

which implies

$$|\nabla E|(x_t) = \varlimsup_{y \to x_t} \frac{\big(E(x_t) - E(y)\big)^+}{d(x_t, y)} \le |\dot{x}_t|, \qquad a.e.\ t > 0. \tag{3.10}$$

Fix an interval $[a, b] \subset (0, \infty)$, let L be the Lipschitz constant of (x_t) in $[a, b]$ and observe that for any $y \in X$ it holds

$$\frac{d}{dt}d^2(x_t, y) \ge -|\dot{x}_t| d(x_t, y) \ge -L d(x_t, y), \qquad a.e.\ t \in [a, b].$$

Plugging this bound in the EVI we get

$$-L d(x_t, y) + \frac{\lambda}{2}d^2(x_t, y) + E(x_t) \le E(y), \qquad a.e.\ t \in [a, b],$$

and by the lower semicontinuity of $t \mapsto E(x_t)$ the inequality holds for every $t \in [a,b]$. Taking $y = x_s$ and then exchanging the roles of x_t, x_s we deduce

$$\big|E(x_t) - E(x_s)\big| \le Ld(x_t,x_s) - \frac{\lambda}{2}d^2(x_t,x_s) \le L|t-s|\left(L + \frac{|\lambda|}{2}L|t-s|\right), \qquad \forall t,s \in [a,b],$$

thus the map $t \mapsto E(x_t)$ is locally Lipschitz. It is then obvious that it holds

$$\begin{aligned}-\frac{d}{dt}E(x_t) &= \lim_{h\to 0}\frac{E(x_t) - E(x_{t+h})}{h} = \lim_{h\to 0}\frac{E(x_t) - E(x_{t+h})}{d(x_{t+h},x_t)}\frac{d(x_{t+h},x_t)}{h}\\ &\le |\nabla E|(x_t)|\dot{x}_t| \le \frac{1}{2}|\nabla E|^2(x_t) + \frac{1}{2}|\dot{x}_t|^2, \qquad a.e.\ t.\end{aligned}$$

Thus to conclude we need only to prove the opposite inequality. Integrate the EVI from t to $t+h$ to get

$$\frac{d^2(x_{t+h},y) - d^2(x_t,y)}{2} + \int_t^{t+h} E(x_s)\,ds + \int_t^{t+h}\frac{\lambda}{2}d^2(x_s,y)ds \le hE(y).$$

Let $y = x_t$ to obtain

$$\frac{d^2(x_{t+h},x_t)}{2} \le \int_t^{t+h} E(x_t) - E(x_s)\,ds + \frac{|\lambda|}{6}L^2h^3 = h\int_0^1 E(x_t) - E(x_{t+hr})\,dr + \frac{|\lambda|}{6}L^2h^3.$$

Now let $A \subset (0,+\infty)$ be the set of points of differentiability of $t \mapsto E(x_t)$ and where $|\dot{x}_t|$ exists, choose $t \in A \cap (a,b)$, divide by h^2 the above inequality, let $h \to 0$ and use the dominated convergence theorem to get

$$\frac{1}{2}|\dot{x}_t|^2 \le \lim_{h\to 0}\int_0^1 \frac{E(x_t) - E(x_{t+hr})}{h}dr = -\frac{d}{dt}E(x_t)\int_0^1 r\,dr = -\frac{1}{2}\frac{d}{dt}E(x_t).$$

Recalling (3.10) we conclude with

$$-\frac{d}{dt}E(x_t) \ge |\dot{x}_t|^2 \ge \frac{1}{2}|\dot{x}_t|^2 + \frac{1}{2}|\nabla E|^2(x_t), \qquad a.e.\ t > 0.$$

Finally, we see how the local Lipschitz property of (x_t) can be achieved. It is immediate to verify that the curve $t \mapsto x_{t+h}$ is a Gradient Flow in the EVI sense starting from x_h for all $h > 0$. We now use the fact that the distance between curves satisfying the EVI is contractive up to an exponential factor (see the last part of the proof of Theorem 3.25 for a sketch of the argument, and Corollary 4.3.3 of [6] for the rigorous proof). We have

$$d(x_s,x_{s+h}) \le e^{-\lambda(s-t)}d(x_t,x_{t+h}), \qquad \forall s > t.$$

Dividing by h, letting $h \downarrow 0$ and calling $B \subset (0,\infty)$ the set where the metric derivative of x_t exists, we obtain

$$|\dot{x}_s| \le |\dot{x}_t|e^{-\lambda(s-t)}, \qquad \forall s,\, t \in B,\ s > t.$$

This implies that the curve (x_t) is locally Lipschitz in $(0,+\infty)$. □

Let us come back to the case of a convex and lower semicontinuous functional F on an Hilbert space. Pick $\overline{x} \in \overline{D(F)}$, fix $\tau > 0$ and define the sequence $n \mapsto x^\tau_{(n)}$ recursively by setting $x^\tau_{(n)} := \overline{x}$ and defining $x^\tau_{(n+1)}$ as a minimizer of

$$x \qquad \mapsto \qquad F(x) + \frac{|x - x^\tau_{(n)}|^2}{2\tau}.$$

It is immediate to verify that a minimum exists and that it is unique, thus the sequence $n \mapsto x^\tau_{(n)}$ is well defined. The Euler-Lagrange equation of $x^\tau_{(n+1)}$ is:

$$\frac{x^\tau_{(n+1)} - x^\tau_{(n)}}{\tau} \in -\partial^- F(x^\tau_{(n+1)}),$$

which is a time discretization of (3.2). It is then natural to introduce the rescaled curve $t \mapsto x^\tau_t$ by

$$x^\tau_t := x^\tau_{([t/\tau])},$$

where $[\cdot]$ denotes the integer part, and to ask whether the curves $t \mapsto x^\tau_t$ converge in some sense to a limit curve (x_t) which solves (3.2) as $\tau \downarrow 0$. This is the case, and this procedure is actually the heart of the proof of Theorem 3.1.

What is important for the discussion we are making now, is that the minimization procedure just described can be naturally posed in a metric setting for a general functional $E : X \to \mathbb{R} \cup \{+\infty\}$: it is sufficient to pick $\overline{x} \in \overline{\{E < \infty\}}$, $\tau > 0$, define $x^\tau_{(0)} := \overline{x}$ and then recursively

$$x^\tau_{(n+1)} \in \operatorname{argmin} \left\{ x \mapsto E(x) + \frac{d^2(x, x^\tau_{(n)})}{2\tau} \right\}. \tag{3.11}$$

We this give the following definition:

Definition 3.7 (Discrete solution) *Let (X, d) be a metric space, $E : X \to \mathbb{R} \cup \{+\infty\}$ lower semi-continuous, $\overline{x} \in \overline{\{E < \infty\}}$ and $\tau > 0$. A* discrete solution *is a map $[0, +\infty) \ni t \mapsto x^\tau_t$ defined by*

$$x^\tau_t := x^\tau_{([t/\tau])},$$

where $x^\tau_{(0)} := \overline{x}$ and $x^\tau_{(n+1)}$ satisfies (3.11).

Clearly in a metric context it is part of the job the identification of suitable assumptions that ensure that the minimization problem (3.11) admits at least a minimum, so that discrete solutions exist.

We now divide the discussion into three parts, to see under which conditions on the functional E and the metric space X it is possible to prove existence of Gradient Flows in the EDI, EDE and EVI formulation.

3.2.2 General l.s.c. functionals and EDI

In this section we will make minimal assumptions on the functional E and show how it is possible, starting from them, to prove existence of Gradient Flows in the EDI sense.

Basically, there are two "independent" sets of assumptions that we need: those which ensure the existence of discrete solutions, and those needed to pass to the limit. To better highlight the structure of the theory, we first introduce the hypotheses we need to guarantee the existence of discrete solution and see which properties the discrete solutions have. Then, later on, we introduce the assumptions needed to pass to the limit.

We will denote by $D(E) \subset X$ the domain of E, i.e. $D(E) := \{E < \infty\}$

Assumption 3.8 (Hypothesis for existence of discrete solutions) *(X, d) is a Polish space and $E : X \to \mathbb{R} \cup \{+\infty\}$ be a l.s.c. functional bounded from below. Also, we assume that there exists $\overline{\tau} > 0$ such that for every $0 < \tau < \overline{\tau}$ and $\overline{x} \in \overline{D(E)}$ there exists at least a minimum of*

$$x \quad \mapsto \quad E(x) + \frac{d^2(x, \overline{x})}{2\tau}. \tag{3.12}$$

Thanks to our assumptions we know that discrete solutions exist for every starting point $\overline{x}$, for τ sufficiently small. The big problem we have to face now is to show that the discrete solutions satisfy a discretized version of the EDI suitable to pass to the limit. The key enabler to do this, is the following result, due to de Giorgi.

Theorem 3.9 (Properties of the variational interpolation) *Let X, E be satisfying the Assumption 3.8. Fix $\overline{x} \in X$, and for any $0 < \tau < \overline{\tau}$ choose x_τ among the minimizers of* (3.12). *Then the map $\tau \mapsto E(x_\tau) + \frac{d^2(\overline{x}, x_\tau)}{2\tau}$ is locally Lipschitz in $(0, \overline{\tau})$ and it holds*

$$\frac{d}{d\tau}\left(E(x_\tau) + \frac{d^2(x, x_\tau)}{2\tau}\right) = -\frac{d^2(x, x_\tau)}{2\tau^2}, \qquad a.e.\ \tau \in (0, \overline{\tau}). \tag{3.13}$$

Proof Observe that from $E(x_{\tau_0}) + \frac{d^2(x_{\tau_0}, x)}{2\tau_0} \leq E(x_{\tau_1}) + \frac{d^2(x_{\tau_1}, x)}{2\tau_0}$ we deduce

$$E(x_{\tau_0}) + \frac{d^2(x_{\tau_0}, x)}{2\tau_0} - E(x_{\tau_1}) + \frac{d^2(x_{\tau_1}, x)}{2\tau_1} \leq \left(\frac{1}{2\tau_0} - \frac{1}{2\tau_1}\right) d^2(x_{\tau_1}, x) = \frac{\tau_1 - \tau_0}{2\tau_0\tau_1} d^2(x_{\tau_1}, x).$$

Arguing symmetrically we see that

$$E(x_{\tau_0}) + \frac{d^2(x_{\tau_0}, x)}{2\tau_0} - E(x_{\tau_1}) + \frac{d^2(x_{\tau_1}, x)}{2\tau_1} \geq \frac{\tau_1 - \tau_0}{2\tau_0\tau_1} d^2(x_{\tau_0}, x).$$

The last two inequalities show that $\tau \mapsto E(x_\tau) + \frac{d^2(x, x_\tau)}{2\tau}$ is locally Lipschitz and that equation (3.13) holds. □

Lemma 3.10 *With the same notation and assumptions as in the previous theorem, $\tau \mapsto d(\overline{x}, x_\tau)$ is non decreasing and $\tau \mapsto E(x_\tau)$ is non increasing. Also, it holds*

$$|\nabla E|(x_\tau) \leq \frac{d(x_\tau, \overline{x})}{\tau}. \tag{3.14}$$

Proof Pick $0 < \tau_0 < \tau_1 < \overline{\tau}$. From the minimality of x_{τ_0} and x_{τ_1} we get

$$\begin{aligned} E(x_{\tau_0}) + \frac{d^2(x_{\tau_0}, \overline{x})}{2\tau_0} &\leq E(x_{\tau_1}) + \frac{d^2(x_{\tau_1}, \overline{x})}{2\tau_0}, \\ E(x_{\tau_1}) + \frac{d^2(x_{\tau_1}, \overline{x})}{2\tau_1} &\leq E(x_{\tau_0}) + \frac{d^2(x_{\tau_0}, \overline{x})}{2\tau_1}. \end{aligned}$$

Adding up and using the fact that $\frac{1}{\tau_0} - \frac{1}{\tau_1} \geq 0$ we get $d(\overline{x}, x_{\tau_0}) \leq d(\overline{x}, x_{\tau_1})$. The fact that $\tau \mapsto E(x_\tau)$ is non increasing now follows from

$$E(x_{\tau_1}) + \frac{d^2(x_{\tau_0}, \overline{x})}{2\tau_1} \leq E(x_{\tau_1}) + \frac{d^2(x_{\tau_1}, \overline{x})}{2\tau_1} \leq E(x_{\tau_0}) + \frac{d^2(x_{\tau_0}, \overline{x})}{2\tau_1}.$$

For the second part of the statement, observe that from

$$E(x_\tau) + \frac{d^2(x_\tau, \overline{x})}{2\tau} \leq E(y) + \frac{d^2(y, \overline{x})}{2\tau}, \qquad \forall y \in X$$

we get

$$\begin{aligned} \frac{E(x_\tau) - E(y)}{d(x_\tau, y)} &\leq \frac{d^2(y, \overline{x}) - d^2(x_\tau, \overline{x})}{2\tau d(x_\tau, y)} = \frac{\big(d(y, \overline{x}) - d(x_\tau, \overline{x})\big)\big(d(x_\tau, \overline{x}) + d(y, \overline{x},)\big)}{2\tau d(x_\tau, y)} \\ &\leq \frac{d(x_\tau, \overline{x},) + d(y, \overline{x})}{2\tau}. \end{aligned}$$

Taking the limsup as $y \to x_\tau$ we get the thesis. □

By Theorem 3.9 and Lemma 3.10 it is natural to introduce the following *variational interpolation* in the Minimizing Movements scheme (as opposed to the classical piecewise constant/affine interpolations used in other contexts):

Definition 3.11 (Variational interpolation) *Let X, E be satisfying Assumption 3.8, $\overline{x} \in \overline{D(E)}$ and $0 < \tau < \overline{\tau}$. We define the map $[0, \infty) \ni t \mapsto x_t^\tau$ in the following way:*

- $x_0^\tau := \overline{x}$,
- $x_{(n+1)\tau}^\tau$ *is chosen among the minimizers of* (3.12) *with $\overline{x}$ replaced by $x_{n\tau}^\tau$,*
- x_t^τ *with $t \in (n\tau, (n+1)\tau)$ is chosen among the minimizers of* (3.12) *with $\overline{x}$ and τ replaced by $x_{n\tau}^\tau$ and $t - n\tau$ respectively.*

For (x_t^τ) defined in this way, we define the discrete speed $\mathrm{Dsp}^\tau : [0, +\infty) \to [0, +\infty)$ *and the* Discrete slope $\mathrm{Dsl}^\tau : [0, +\infty) \to [0, +\infty)$ *by:*

$$\begin{aligned}\mathrm{Dsp}_t^\tau &:= \frac{d\big(x_{n\tau}^\tau, x_{(n+1)\tau}^\tau\big)}{\tau}, \qquad t \in (n\tau, (n+1)\tau),\\ \mathrm{Dsl}_t^\tau &:= \frac{d\big(x_t^\tau, x_{n\tau}^\tau\big)}{t - n\tau}, \qquad t \in (n\tau, (n+1)\tau).\end{aligned} \tag{3.15}$$

Although the object Dsl_t^τ does not look like a slope, we chose this name because from (3.14) we know that $|\nabla E|(x_t^\tau) \leq \mathrm{Dsl}_t^\tau$ and because in the limiting process Dsl^τ will produce the slope term in the EDI (see the proof of Theorem 3.14).

With this notation we have the following result:

Corollary 3.12 (EDE for the discrete solutions) *Let X, E be satisfying Assumption 3.8, $\overline{x} \in \overline{D(E)}$, $0 < \tau < \overline{\tau}$ and (x_t^τ) defined via the variational interpolation as in Definition 3.11 above. Then it holds*

$$E(x_s^\tau) + \frac{1}{2}\int_t^s |\mathrm{Dsp}_r^\tau|^2 dr + \frac{1}{2}\int_t^s |\mathrm{Dsl}_r^\tau|^2 dr = E(x_t^\tau), \tag{3.16}$$

for every $t = n\tau$, $s = m\tau$, $n < m \in \mathbb{N}$.

Proof It is just a restatement of equation (3.13) in terms of the notation given in (3.15). □

Thus, at the level of discrete solutions, it is possible to get a discrete form of the Energy Dissipation Equality under the quite general Assumptions 3.8. Now we want to pass to the limit as $\tau \downarrow 0$. In order to do this, we need to add some compactness and regularity assumptions on the functional:

Assumption 3.13 (Coercivity and regularity assumptions) *Assume that $E : X \to \mathbb{R} \cup \{+\infty\}$ satisfies:*

- *E is bounded from below and its sublevels are boundedly compact, i.e. $\{E \leq c\} \cap \overline{B_r(x)}$ is compact for any $c \in \mathbb{R}$, $r > 0$ and $x \in X$,*
- *the slope $|\nabla E| : D(E) \to [0, +\infty]$ is lower semicontinuous,*
- *E has the following continuity property:*

$$x_n \to x,\ \sup_n\{|\nabla E|(x_n), E(x_n)\} < \infty \qquad \Rightarrow \qquad E(x_n) \to E(x).$$

Under these assumptions we can prove the following result:

Theorem 3.14 (Gradient Flows in EDI formulation) *Let (X, d) be a metric space and let $E : X \to \mathbb{R} \cup \{+\infty\}$ be satisfying the Assumptions 3.8 and 3.13. Also, let $\overline{x} \in D(E)$ and for $0 < \tau < \overline{\tau}$ define the discrete solution via the variational interpolation as in Definition 3.11. Then it holds:*

- *the set of curves $\{(x_t^\tau)\}_\tau$ is relatively compact in the set of curves in X w.r.t. local uniform convergence,*
- *any limit curve (x_t) is a Gradient Flow in the EDI formulation (Definition 3.3).*

Sketch of the Proof
Compactness. By Corollary 3.12 we have

$$d^2(x_t^\tau, \overline{x}) \le \left(\int_0^T |\mathrm{Dsp}_r^\tau| dr \right)^2 \le T \int_0^T |\mathrm{Dsp}_\mathrm{r}^\tau|^2 dr \le 2T\big(E(\overline{x}) - \inf E\big), \qquad \forall t \le T,$$

for any $T = n\tau$. Therefore for any $T > 0$ the set $\{x_t^\tau\}_{t \le T}$ is uniformly bounded in τ. As this set is also contained in $\{E \le E(\overline{x})\}$, it is relatively compact. The fact that there is relative compactness w.r.t. local uniform convergence follows by an Ascoli-Arzelà-type argument based on the inequality

$$d^2\big(x_t^\tau, x_s^\tau\big) \le \left(\int_t^s |\mathrm{Dsp}_r^\tau| dr \right)^2 \le 2(s-t)\big(E(\overline{x}) - \inf E\big), \qquad \forall t = n\tau,\ s = m\tau,\ n < m \in \mathbb{N}. \tag{3.17}$$

Passage to the limit. Let $\tau_n \downarrow 0$ be such that $(x_t^{\tau_n})$ converges to a limit curve x_t locally uniformly. Then by standard arguments based on inequality (3.17) it is possible to check that $t \mapsto x_t$ is absolutely continuous and satisfies

$$\int_t^s |\dot{x}_r|^2 dr \le \varliminf_{n\to\infty} \int_t^s |\mathrm{Dsp}_r^{\tau_n}|^2 dr \qquad \forall 0 \le t < s. \tag{3.18}$$

By the lower semicontinuity of $|\nabla E|$ and (3.14) we get

$$|\nabla E|(x_t) \le \varliminf_{n\to\infty} |\nabla E|(x_t^{\tau_n}) \le \varliminf_{n\to\infty} \mathrm{Dsl}_t^{\tau_n}, \qquad \forall t,$$

thus Fatou's lemma ensures that for any $t < s$ it holds

$$\int_t^s |\nabla E|^2(x_r) dr \le \int_t^s \varliminf_{n\to\infty} |\nabla E|^2(x_r^\tau) dr \le \varliminf_{n\to\infty} \int_t^s |\mathrm{Dsl}_r^{\tau_n}|^2\, dr \le 2T\big(E(\overline{x}) - \inf E\big). \tag{3.19}$$

Now passing to the limit in (3.16) written for $t = 0$ we get the first inequality in (3.8). Also, from (3.19) we get that the L^2 norm of $f(t) := \varliminf_{n\to\infty} |\nabla E|(x_t^{\tau_n})$ on $[0,\infty)$ is finite. Thus $A := \{f < \infty\}$ has full Lebesgue measure and for each $t \in A$ we can find a subsequence $\tau_{n_k} \downarrow 0$ such that $\sup_k |\nabla E|(x_t^{\tau_{n_k}}) < \infty$. Then the third assumption in 3.13 guarantees that $E(x_t^{\tau_{n_k}}) \to E(x_t)$ and the lower semicontinuity of E that $E(x_s) \le \varliminf_{k\to\infty} E(x_s^{\tau_{n_k}})$ for every $s \ge t$. Thus passing to the limit in (3.16) as $\tau_{n_k} \downarrow 0$ and using (3.18) and (3.19) we get

$$E(x_s) + \frac{1}{2}\int_t^s |\dot{x}_r|^2 dr + \frac{1}{2}\int_t^s |\nabla E|^2(x_r) dr \le E(x_t), \qquad \forall t \in A,\ \forall s \ge t.$$

□

We conclude with an example which shows why in general we cannot hope to have equality in the EDI. Shortly said, the problem is that we don't know whether $t \mapsto E(x_t)$ is an absolutely continuous map.

Example 3.15 Let $X = [0,1]$ with the Euclidean distance, $C \subset X$ a Cantor-type set with null Lebesgue measure and $f : [0,1] \to [1,+\infty]$ a continuous, integrable function such that $f(x) = +\infty$

for any $x \in C$, which is smooth on the complement of C. Also, let $g : [0,1] \to [0,1]$ be a "Devil staircase" built over C, i.e. a continuous, non decreasing function satisfying $g(0) = 0$, $g(1) = 1$ which is constant in each of the connected components of the complement of C. Define the energies $E, \tilde{E} : [0,1] \to \mathbb{R}$ by

$$E(x) := -g(x) - \int_0^x f(y)dy.$$
$$\tilde{E}(x) := -\int_0^x f(y)dy.$$

It is immediate to verify that $E, \tilde{E}$ satisfy all the Assumptions 3.8, 3.13 (the choice of f guarantees that the slopes of $E, \tilde{E}$ are continuous). Now build a Gradient Flow starting from 0: with some work it is possible to check that the Minimizing Movement scheme converges in both cases to absolutely continuous curves (x_t) and $(\tilde{x}_t)$ respectively satisfying

$$x_t' = -|\nabla E|(x_t), \qquad a.e.\ t$$
$$\tilde{x}_t' = -|\nabla \tilde{E}|(\tilde{x}_t), \qquad a.e.\ t.$$

Now, notice that $|\nabla E|(x) = |\nabla \tilde{E}|(x) = f(x)$ for every $x \in [0,1]$, therefore the fact that $f \geq 1$ is smooth on $[0,1] \setminus C$ gives that each of these two equations admit a unique solution. Therefore - this is the key point of the example - (x_t) and $(\tilde{x}_t)$ must coincide. In other words, the effect of the function g is not seen at the level of Gradient Flow. It is then immediate to verify that there is Energy Dissipation Equality for the energy $\tilde{E}$, but there is only the Energy Dissipation Inequality for the energy E. ■

3.2.3 The geodesically convex case: EDE and regularizing effects

Here we study gradient flows of so called *geodesically convex* functionals, which are the natural metric generalization of convex functionals on linear spaces.

Definition 3.16 (Geodesic convexity) *Let $E : X \to \mathbb{R} \cup \{+\infty\}$ be a functional and $\lambda \in \mathbb{R}$. We say that E is λ-geodesically convex provided for every x, $y \in X$ there exists a constant speed geodesic $\gamma : [0,1] \to X$ connecting x to y such that*

$$E(\gamma_t) \leq (1-t)E(x) + tE(y) - \frac{\lambda}{2}t(1-t)d^2(x,y). \tag{3.20}$$

In this section we will assume that:

Assumption 3.17 (Geodesic convexity hypothesis) (X, d) *is a Polish geodesic space, $E : X \to \mathbb{R} \cup \{+\infty\}$ is lower semicontinuous, λ-geodesically convex for some $\lambda \in \mathbb{R}$. Also, we assume that the sublevels of E are boundedly compact, i.e. the set $\{E \leq c\} \cap \overline{B_r(x)}$ is compact for any $c \in \mathbb{R}$, $r > 0$, $x \in X$.*

What we want to prove is that for X, E satisfying these assumptions there is existence of Gradient Flows in the formulation EDE (Definition 3.4).

Our first goal is to show that in this setting it is possible to recover the results of the previous section. We start claiming that it holds:

$$|\nabla E|(x) = \sup_{y \neq x} \left(\frac{E(x) - E(y)}{d(x,y)} + \frac{\lambda}{2} d(x,y) \right)^+, \tag{3.21}$$

so that the $\overline{\lim}$ in the definition of the slope can be replaced by a $\sup$. Indeed, we know that

$$|\nabla E|(x) = \overline{\lim_{y\to x}}\left(\frac{E(x)-E(y)}{d(x,y)} + \frac{\lambda}{2}d(x,y)\right)^+ \leq \sup_{y\neq x}\left(\frac{E(x)-E(y)}{d(x,y)} + \frac{\lambda}{2}d(x,y)\right)^+.$$

To prove the opposite inequality fix $y \neq x$ and a constant speed geodesic γ connecting x to y for which (3.20) holds. Then observe that

$$\begin{aligned}|\nabla E|(x) &\geq \overline{\lim_{t\downarrow 0}}\left(\frac{E(x)-E(\gamma_t)}{d(x,\gamma_t)}\right)^+ = \left(\overline{\lim_{t\downarrow 0}}\,\frac{E(x)-E(\gamma_t)}{d(x,\gamma_t)}\right)^+ \\ &\stackrel{(3.20)}{\geq} \left(\overline{\lim_{t\downarrow 0}}\left(\frac{E(x)-E(y)}{d(x,y)} + \frac{\lambda}{2}(1-t)d(x,y)\right)\right)^+ = \left(\frac{E(x)-E(y)}{d(x,y)} + \frac{\lambda}{2}d(x,y)\right)^+.\end{aligned}$$

Using this representation formula we can show that all the assumptions 3.8 and 3.13 hold:

Proposition 3.18 *Suppose that Assumption 3.17 holds. Then Assumptions 3.8 and 3.13 hold as well.*

Sketch of the Proof From the λ-geodesic convexity and the lower semicontinuity assumption it is possible to deduce (we omit the details) that E has at most quadratic decay at infinity, i.e. there exists $\overline{x} \in X$, $a, b > 0$ such that

$$E(x) \geq -a - bd(x,\overline{x}) + \lambda^- d^2(x,\overline{x}), \qquad \forall x \in X.$$

Therefore from the lower semicontinuity again and the bounded compactness of the sublevels of E we immediately get that the minimization problem (3.12) admits a solution if $\tau < 1/\lambda^-$.

The lower semicontinuity of the slope is a direct consequence of (3.21) and of the lower semicontinuity of E. Thus, to conclude we need only to show that

$$x_n \to x,\ \sup_n\{|\nabla E|(x_n), E(x_n)\} < \infty \qquad \Rightarrow \qquad \overline{\lim_{n\to\infty}}\, E(x_n) \leq E(x). \tag{3.22}$$

From (3.21) with x, y replaced by x_n, x respectively we get

$$E(x) \geq E(x_n) - |\nabla E|(x_n)d(x,x_n) + \frac{\lambda}{2}d^2(x,x_n),$$

and the conclusion follows by letting $n \to \infty$. $\square$

Thus Theorem 3.14 applies directly also to this case and we get existence of Gradient Flows in the EDI formulation. To get existence in the stronger EDE formulation, we need the following result, which may be thought as a sort of weak chain rule (observe that the validity of the proposition below rules out behaviors like the one described in Example 3.15).

Proposition 3.19 *Let E be a λ-geodesically convex and l.s.c. functional. Then for every absolutely continuous curve $(x_t) \subset X$ such that $E(x_t) < \infty$ for every t, it holds*

$$\big|E(x_s) - E(x_t)\big| \leq \int_t^s |\dot{x}_r||\nabla E(x_r)|dr, \qquad \forall t < s. \tag{3.23}$$

Proof We may assume that the right hand side of (3.23) is finite for any $t, s \in [0,1]$, and, by a reparametrization argument, we may also assume that $|\dot{x_t}| = 1$ for a.e. t (in particular (x_t) is

1-Lipschitz), so that $t \mapsto |\nabla E|(x_t)$ is an L^1 function. Notice that it is sufficient to prove that $t \mapsto E(x_t)$ is absolutely continuous, as then the inequality

$$\begin{aligned}
\overline{\lim_{h\uparrow 0}} \frac{E(x_{t+h}) - E(x_t)}{h} &\leq \overline{\lim_{h\uparrow 0}} \frac{(E(x_t) - E(x_{t+h}))^+}{|h|} \\
&\leq \overline{\lim_{h\uparrow 0}} \frac{(E(x_t) - E(x_{t+h}))^+}{d(x_t, x_{t+h})} \overline{\lim_{h\uparrow 0}} \frac{d(x_t, x_{t+h})}{|h|} \leq |\nabla E(x_t)||\dot{x}_t|,
\end{aligned}$$

valid for any $t \in [0,1]$ gives (3.23).

Define the functions $f, g : [0,1] \to \mathbb{R}$ by

$$\begin{aligned}
f(t) &:= E(x_t), \\
g(t) &:= \sup_{s\neq t} \frac{(f(t) - f(s))^+}{|s-t|}
\end{aligned}$$

Let D be the diameter of the compact set $\{x_t\}_{t\in[0,1]}$, use the fact that (x_t) is 1-Lipschitz, formula (3.21) and the trivial inequality $a^+ \leq (a+b)^+ + b^-$ (valid for any a, $b \in \mathbb{R}$) to get

$$g(t) \leq \sup_{s\neq t} \frac{(E(x_t) - E(x_s))^+}{d(x_s, x_t)} \leq |\nabla E|(x_t) + \frac{\lambda^-}{2} D.$$

Therefore the thesis will be proved if we show that:

$$g \in L^1 \qquad \Rightarrow \qquad |f(s) - f(t)| \leq \int_t^s g(r)dr \qquad \forall t < s. \tag{3.24}$$

Fix $M > 0$ and define $f^M := \min\{f, M\}$. Now fix $\varepsilon > 0$, pick a smooth mollifier $\rho_\varepsilon : \mathbb{R} \to \mathbb{R}$ with support in $[-\varepsilon, \varepsilon]$ and define f^M_ε, $g^M_\varepsilon : [\varepsilon, 1-\varepsilon] \to \mathbb{R}$ by

$$\begin{aligned}
f^M_\varepsilon(t) &:= f^M * \rho_\varepsilon(t), \\
g^M_\varepsilon(t) &:= \sup_{s\neq t} \frac{(f^M_\varepsilon(t) - f^M_\varepsilon(s))^+}{|s-t|}.
\end{aligned}$$

Since f^M_ε is smooth and $g^M_\varepsilon \geq (f^M_\varepsilon)'$ it holds

$$|f^M_\varepsilon(s) - f^M_\varepsilon(t)| \leq \int_t^s g^M_\varepsilon(r)dr. \tag{3.25}$$

From the trivial bound $(\int h)^+ \leq \int h^+$ we get

$$\begin{aligned}
g^M_\varepsilon(t) &\leq \sup_s \frac{\int (f^M(t-r) - f^M(s-r))^+ \rho_\varepsilon(r)dr}{|s-t|} \leq \sup_s \frac{\int (f(t-r) - f(s-r))^+ \rho_\varepsilon(r)dr}{|s-t|} \\
&= \sup_s \int \frac{(f(t-r) - f(s-r))^+}{|(s-r) - (t-r)|} \rho_\varepsilon(r)dr \leq \int g(t-r)\rho_\varepsilon(r)dr = g * \rho_\varepsilon(t).
\end{aligned} \tag{3.26}$$

Thus the family of functions $\{g^M_\varepsilon\}_\varepsilon$ is dominated in $L^1(0,1)$. From (3.25) and (3.26) it follows that the family of functions $\{f^M_\varepsilon\}$ uniformly converge to some function $\tilde{f}^M$ on $[0,1]$ as $\varepsilon \downarrow 0$ for which it holds

$$|\tilde{f}^M(s) - \tilde{f}^M(t)| \leq \int_t^s g(r)dr.$$

We know that $f^M = \tilde{f}^M$ on some set $A \subset [0,1]$ such that $\mathcal{L}^1([0,1] \setminus A) = 0$, and we want to prove that they actually coincide everywhere. Recall that f^M is l.s.c. and $\tilde{f}^M$ is continuous, hence $f^M \leq \tilde{f}^M$ in $[0,1]$. If by contradiction it holds $f^M(t_0) < c < C < \tilde{f}^M(t_0)$ for some t_0, c, C, we can find $\delta > 0$ such that $\tilde{f}^M(t) > C$ in $t \in [t_0-\delta, t_0+\delta]$. Thus $f^M(t) > C$ for $t \in [t_0-\delta, t_0+\delta] \cap A$ and the contradiction comes from

$$\int_0^1 g(t)dt \geq \int_{[t_0-\delta, t_0+\delta]\cap A} g(t)dt \geq \int_{[t_0-\delta, t_0+\delta]\cap A} \frac{C-c}{|t-t_0|} dt = +\infty.$$

Thus we proved that if $g \in L^1(0,1)$ it holds

$$|f^M(t) - f^M(s)| \leq \int_t^s g(r)dr, \qquad \forall t < s \in [0,1],\ M > 0.$$

Letting $M \to \infty$ we prove (3.24) and hence the thesis. $\square$

This proposition is the key ingredient to pass from existence of Gradient Flows in the EDI formulation to the one in the EDE formulation:

Theorem 3.20 (Gradient Flows in the EDE formulation) *Let X, E be satisfying Assumption 3.17 and $\overline{x} \in X$ be such that $E(\overline{x}) < \infty$. Then all the results of Theorem 3.14 hold.*

Also, any Gradient Flow in the EDI sense is also a Gradient Flow in the EDE sense (Definition 3.4).

Proof The first part of the statement follows directly from Proposition 3.18.

By Theorem 3.14 we know that the limit curve is absolutely continuous and satisfies

$$E(x_s) + \frac{1}{2}\int_0^s |\dot{x}|_r^2 dr + \frac{1}{2}\int_0^s |\nabla E|^2(x_r)dr \leq E(\overline{x}), \qquad \forall s \geq 0. \tag{3.27}$$

In particular, the functions $t \mapsto |\dot{x}_t|$ and $t \mapsto |\nabla E|(x_t)$ belong to $L^2_{loc}(0,+\infty)$. Now we use Proposition 3.19: we know that for any $s \geq 0$ it holds

$$\big|E(\overline{x}) - E(x_s)\big| \leq \int_0^s |\dot{x}_r||\nabla E|(x_r)dr \leq \frac{1}{2}\int_0^s |\dot{x}_r|^2 dr + \frac{1}{2}\int_0^s |\nabla E|^2(x_r)dr. \tag{3.28}$$

Therefore $t \mapsto E(x_t)$ is locally absolutely continuous and it holds

$$E(x_s) + \frac{1}{2}\int_0^s |\dot{x}_r|^2 dr + \frac{1}{2}\int_0^s |\nabla E|^2(x_r)dr = E(\overline{x}), \qquad \forall s \geq 0.$$

Subtracting from this last equation the same equality written for $s = t$ we get the thesis. $\square$

Remark 3.21 It is important to underline that the hypothesis of λ-geodesic convexity is in general of no help for what concerns the compactness of the sequence of discrete solutions. $\blacksquare$

The λ-geodesic convexity hypothesis, ensures various regularity results for the limit curve, which we state without proof:

Proposition 3.22 *Let X, E be satisfying Assumption 3.17 and let (x_t) be any limit of a sequence of discrete solutions. Then:*

i) the limit

$$|\dot{x}_t^+| := \lim_{h\downarrow 0} \frac{d(x_{t+h}, x_t)}{h},$$

exists for every $t > 0$,

ii) the equation

$$\frac{d}{dt_+} E(x_t) = -|\nabla E|^2(x_t) = -|\dot{x}_t^+|^2 = -|\dot{x}_t^+||\nabla E|(x_t),$$

is satisfied at every $t > 0$,

iii) the map $t \mapsto e^{-2\lambda^- t} E(x_t)$ is convex, the map $t \mapsto e^{\lambda t}|\nabla E|(x_t)$ is non increasing, right continuous and satisfies

$$\frac{t}{2}|\nabla E|^2(x_t) \leq e^{2\lambda^- t}\Big(E(x_0) - E_t(x_0)\Big),$$
$$t|\nabla E|^2(x_t) \leq (1 + 2\lambda^+ t)e^{-2\lambda t}\Big(E(x_0 - \inf E\Big),$$

where $E_t : X \to \mathbb{R}$ is defined as

$$E_t(x) := \inf_y E(y) + \frac{d^2(x,y)}{2t},$$

iv) if $\lambda > 0$, then E admits a unique minimum x_{min} and it holds

$$\frac{\lambda}{2} d^2(x_t, x_{min}) \leq E(x_t) - E(x_{min}) \leq e^{-2\lambda t}\big(E(x_0) - E(x_{min})\big).$$

Observe that we didn't state any result concerning the uniqueness (nor about contractivity) of the curve (x_t) satisfying the Energy Dissipation Equality (3.9). The reason is that if no further assumptions are made on either X or E, in general uniqueness fails, as the following simple example shows:

Example 3.23 (Lack of uniqueness) Let $X := \mathbb{R}^2$ endowed with the L^∞ norm, $E : X \to \mathbb{R}$ be defined by $E(x^1, x^2) := x^1$ and $\overline{x} := (0,0)$. Then it is immediate to verify that $|\nabla E| \equiv 1$ and that any Lipschitz curve $t \mapsto x_t = (x_t^1, x_t^2)$ satisfying

$$x_t^1 = -t, \qquad \forall t \geq 0$$
$$|x_t^{2'}| \leq 1, \qquad a.e.\ t > 0,$$

satisfies also

$$E(x_t) = -t,$$
$$|\dot{x}_t| = 1.$$

This implies that any such (x_t) satisfies the Energy Dissipation Equality (3.9). ■

3.2.4 The compatibility of Energy and distance: EVI and error estimates

As the last example of the previous section shows, in general we cannot hope to have uniqueness of the limit curve (x_t) obtained via the Minimizing Movements scheme for a generic λ-geodesically convex functional. If we want to derive properties like uniqueness and contractivity of the flow, we need to have some stronger relation between the Energy functional E and the distance d on X: in this section we will assume the following:

Assumption 3.24 (Compatibility in Energy and distance) *(X, d) is a Polish space. $E : X \to \mathbb{R} \cup \{+\infty\}$ is a lower semicontinuous functional and for any x_0, x_1, $y \in X$, there exists a curve $t \mapsto \gamma(t)$ such that*

$$\begin{aligned} E(\gamma_t) &\leq (1-t)E(x_0) + tE(x_1) - \frac{\lambda}{2}t(1-t)d^2(x_0, x_1), \\ d^2(\gamma_t, y) &\leq (1-t)d^2(x_0, y) + td^2(x_1, y) - t(1-t)d^2(x_0, x_1), \end{aligned} \tag{3.29}$$

for every $t \in [0, 1]$.

Observe that there is no compactness assumption of the sublevels of E. If X is an Hilbert space (and more generally a NPC space - Definition 2.19) then the second inequality in (3.29) is satisfied by geodesics. Hence λ-convex functionals are automatically compatible with the metric.

Following the same lines of the previous section, it is possible to show that this assumption implies both Assumption 3.8 and, if the sublevels of E are boundedly compact, Assumption 3.13, so that Theorem 3.14 holds. Also it can be shown that formula (3.21) is true and thus that Proposition 3.19 holds also in this setting, so that Theorem 3.20 can be proved as well.

However, if Assumption 3.24 holds, it is better not to follow the general theory as developed before, but to restart from scratch: indeed, in this situation much stronger statements hold, also at the level of discrete solutions, which can be proved by a direct use of Assumption 3.24.

We collect the main results achievable in this setting in the following theorem:

Theorem 3.25 (Gradient Flows for compatible E and d: EVI) *Assume that X, E satisfy Assumption 3.24. Then the following hold.*

- *For every $x \in \overline{D(E)}$ and $0 < \tau < 1/\lambda^-$ there exists a unique discrete solution (x_t^τ) as in Definition 3.7.*
- *Let $x \in \overline{D(E)}$ and (x_t^τ) any family of discrete solutions starting from it. Then (x_t^τ) converge locally uniformly to a limit curve (x_t) as $\tau \downarrow 0$ (so that the limit curve is unique). Furthermore, (x_t) is the unique solution of the system of differential inequalities:*

$$\frac{1}{2}\frac{d}{dt}d^2(\tilde{x}_t, y) + \frac{\lambda}{2}d^2(\tilde{x}_t, y) + E(\tilde{x}_t) \leq E(y), \qquad a.e.\ t \geq 0,\ \forall y \in X, \tag{3.30}$$

  among all locally absolutely continuous curves $(\tilde{x}_t)$ converging to $\overline{x}$ as $t \downarrow 0$. I.e. x_t is a Gradient Flow in the EVI formulation - see Definition 3.5.
- *Let $\overline{x}$, $\overline{y} \in \overline{D(E)}$ and (x_t), (y_t) be the two Gradient Flows in the EVI formulation. Then there is λ-exponential contraction of the distance, i.e.:*

$$d^2(x_t, y_t) \leq e^{-\lambda t}d^2(\overline{x}, \overline{y}). \tag{3.31}$$

- *Suppose that $\lambda \geq 0$, that $\overline{x} \in D(E)$ and build x_t^τ, x_t as above. Then the following a priori error estimate holds:*

$$\sup_{t \geq 0} d(x_t, x_t^\tau) \leq 8\sqrt{\tau(E(\overline{x}) - E(x_t))}. \tag{3.32}$$

Sketch of the Proof We will make the following simplifying assumptions: $E \geq 0$, $\lambda \geq 0$ and $\overline{x} \in D(E)$. Also we will prove just that the sequence of discrete solutions $n \mapsto x_t^{\tau/2^n}$ converges to a limit curve as $n \to \infty$ for any given $\tau > 0$.

Existence and uniqueness of the discrete solution. Pick $x \in X$. We have to prove that there exists a unique minimizer of (3.12). Let $I \geq 0$ be the infimum of (3.12). Let (x_n) be a minimizing sequence for (3.12), fix n, $m \in \mathbb{N}$ and let $\gamma : [0,1] \to X$ be a curve satisfying (3.29) for $x_0 := x_n$, $x_1 := x_m$ and $y := x$. Using the inequalities (3.29) at $t = 1/2$ we get

$$\begin{aligned} I &\le E(\gamma_{1/2}) + \frac{d^2(\gamma_{1/2}, x)}{2\tau} \\ &\le \frac{1}{2}\left(E(x_n) + \frac{d^2(x_n, x)}{2\tau} + E(x_m) + \frac{d^2(x_m, x)}{2\tau}\right) - \frac{1+\lambda\tau}{8\tau} d^2(x_n, x_m). \end{aligned}$$

Therefore

$$\varlimsup_{n,m\to\infty} \frac{1+\lambda\tau}{8\tau} d^2(x_n, x_m) \le \varlimsup_{n,m\to\infty} \frac{1}{2}\left(E(x_n) + \frac{d^2(x_n, x)}{2\tau} + E(x_m) + \frac{d^2(x_m, x)}{2\tau}\right) - I = 0,$$

and thus the sequence (x_n) is a Cauchy sequence as soon as $0 < \tau < 1/\lambda^-$. This shows uniqueness, existence follows by the l.s.c. of E.

One step estimates We claim that the following discrete version of the EVI (3.30) holds: for any $x \in X$,

$$\frac{d^2(x^\tau, y) - d^2(x, y)}{2\tau} + \frac{\lambda}{2} d^2(x^\tau, y) \le E(y) - E(x^\tau), \qquad \forall y \in X, \tag{3.33}$$

where x^τ is the minimizer of (3.12). Indeed, pick a curve γ satisfying (3.29) for $x_0 := x^\tau$, $x_1 := y$ and $y := x$ and use the minimality of x^τ to get

$$\begin{aligned} E(x^\tau) + \frac{d^2(x, x^\tau)}{2\tau} \le E(\gamma_t) + \frac{d^2(x, \gamma_t)}{2\tau} &\le (1-t)E(x^\tau) + tE(y) - \frac{\lambda}{2} t(1-t) d^2(x^\tau, y) \\ &\quad + \frac{(1-t)d^2(x, x^\tau) + td^2(x, y) - t(1-t)d^2(x^\tau, y)}{2\tau}. \end{aligned}$$

Rearranging the terms, dropping the positive addend $td^2(x, x^\tau)$ and dividing by $t > 0$ we get

$$\frac{(1-t)d^2(x^\tau, y)}{2\tau} - \frac{d^2(x, y)}{2\tau} + \frac{\lambda}{2}(1-t) d^2(x^\tau, y) \le E(y) - E(x^\tau),$$

so that letting $t \downarrow 0$ we get (3.33).

Now we pass to the discrete version of the error estimate, which will also give the full convergence of the discrete solutions to the limit curve. Given $\overline{x}$, $\overline{y} \in D(E)$, and the associate discrete solutions x_t^τ, y_t^τ, we are going to bound the distance $d(x_\tau^{\tau/2}, y_\tau^\tau)$ in terms of the distance $d(\overline{x}, \overline{y})$.

Write two times the discrete EVI (3.33) for $\tau := \tau/2$ and $y := \overline{y}$: first with $x := \overline{x}$, then with $x := x_{\tau/2}^{\tau/2}$ to get (we use the assumption $\lambda \ge 0$)

$$\begin{aligned} \frac{d^2(x_{\tau/2}^{\tau/2}, \overline{y}) - d^2(\overline{x}, \overline{y})}{\tau} &\le E(\overline{y}) - E(x_{\tau/2}^{\tau/2}), \\ \frac{d^2(x_\tau^{\tau/2}, \overline{y}) - d^2(x_{\tau/2}^{\tau/2}, \overline{y})}{\tau} &\le E(\overline{y}) - E(x_\tau^{\tau/2}). \end{aligned}$$

Adding up these two inequalities and observing that $E(x_\tau^{\tau/2}) \le E(x_{\tau/2}^{\tau/2})$ we obtain

$$\frac{d^2(x_\tau^{\tau/2}, \overline{y}) - d^2(\overline{x}, \overline{y})}{\tau} \le 2\big(E(\overline{y}) - E(x_\tau^{\tau/2})\big).$$

On the other hand, equation (3.33) with $x := \overline{y}$ and $y := x_\tau^{\tau/2}$ reads as

$$\frac{d^2(y_\tau^\tau, x_\tau^{\tau/2}) - d^2(\overline{y}, x_\tau^{\tau/2})}{\tau} \leq 2\big(E(x_\tau^{\tau/2}) - E(y_\tau^\tau)\big).$$

Adding up these last two inequalities we get

$$\frac{d^2(y_\tau^\tau, x_\tau^{\tau/2}) - d^2(\overline{x}, \overline{y})}{\tau} \leq 2\big(E(\overline{y}) - E(y_\tau^\tau)\big). \tag{3.34}$$

Discrete estimates. Pick $t = n\tau < m\tau = s$, write inequality (3.33) for $x := x_{i\tau}^\tau$, $i = n, \ldots, m-1$ and add everything up to get

$$\frac{d^2(x_t^\tau, y) - d^2(x_s^\tau, y)}{2(s-t)} + \frac{\lambda\tau}{2(s-t)} \sum_{i=n+1}^{m} d^2(x_{i\tau}^\tau, y) \leq E(y) - \frac{\tau}{s-t} \sum_{i=n+1}^{m} E(x_{i\tau}^\tau). \tag{3.35}$$

Similarly, pick $t = n\tau$, write inequality (3.34) for $\overline{x} := x_{i\tau}^{\tau/2}$ and $\overline{y} := y_{i\tau}^\tau$ for $i = 0, \ldots, n-1$ and add everything up to get

$$\frac{d^2(x_t^{\tau/2}, y_t^\tau) - d^2(\overline{x}, \overline{y})}{\tau} \leq 2\big(E(\overline{y}) - E(y_t^\tau)\big).$$

Now let $\overline{y} = \overline{x}$ to get

$$d^2(x_t^{\tau/2}, x_t^\tau) \leq 2\tau\big(E(\overline{x}) - E(x_t^\tau)\big) \leq 2\tau E(\overline{x}), \tag{3.36}$$

having used the fact that $E \geq 0$.
Conclusion of passage to the limit. Putting $\tau/2^n$ instead of τ in (3.36) we get

$$d^2(x_t^{\tau/2^{n+1}}, x_t^{\tau/2^n}) \leq \frac{\tau}{2^{n-1}} E(\overline{x}),$$

therefore

$$d^2(x_t^{\tau/2^n}, x_t^{\tau/2^m}) \leq \tau(2^{2-n} - 2^{2-m}) E(\overline{x}), \qquad \forall n < m \in \mathbb{N},$$

which tells that $n \mapsto x_t^{\tau/2^n}$ is a Cauchy sequence for any $t \geq 0$. Also, choosing $n = 0$ and letting $m \to \infty$ we get the error estimate (3.32).

We pass to the EVI. Letting $\tau \downarrow 0$ in (3.35) it is immediate to verify that we get

$$\frac{d^2(x_t, y) - d^2(x_s, y)}{2(s-t)} + \frac{\lambda}{2(s-t)} \int_t^s d^2(x_r, y) \leq E(y) - \frac{1}{s-t} \int_t^s E(x_r) dr,$$

which is precisely the EVI (3.30) written in integral form.
Uniqueness and contractivity. It remains to prove that the solution to the EVI is unique and the contractivity (3.31). The heuristic argument is the following: pick (x_t) and (y_t) solutions of the EVI starting from $\overline{x}$, $\overline{y}$ respectively. Choose $y = y_t$ in the EVI for (x_t) to get

$$\frac{1}{2}\frac{d}{ds}_{|s=t} d^2(x_s, y_t) + \frac{\lambda}{2} d^2(x_t, y_t) + E(x_t) \leq E(y_t).$$

Symmetrically we have

$$\frac{1}{2}\frac{d}{ds}_{|s=t} d^2(x_t, y_s) + \frac{\lambda}{2} d^2(x_t, y_t) + E(y_t) \leq E(x_t).$$

Adding up these two inequalities we get

$$\frac{d}{dt}d^2(x_t, y_t) \leq -2\lambda d^2(x_t, y_t), \qquad a.e.\ t.$$

The rigorous proof follows this line and uses a doubling of variables argument á la Kruzkhov.

Uniqueness and contraction then follow by the Gronwall lemma. $\square$

3.3 Applications to the Wasserstein case

The aim of this section is to apply the abstract theory developed in the previous one to the case of functionals on $(\mathscr{P}_2(\mathbb{R}^d), W_2)$. As we will see, various diffusion equations may be interpreted as Gradient Flows of appropriate energy functionals w.r.t. to the Wasserstein distance, and quantitive analytic properties of the solutions can be derived by this interpretation.

Most of what we are going to discuss here is valid in the more general contexts of Riemannian manifolds and Hilbert spaces, but the differences between these latter cases and the Euclidean one are mainly technical, thus we keep the discussion at a level of $\mathbb{R}^d$ to avoid complications that would just obscure the main ideas.

The secton is split in two subsections: in the first one we discuss the definition of subdifferential of a λ-geodesicaly convex functional on $\mathscr{P}_2(\mathbb{R}^d)$, which is based on the interpretation of $\mathscr{P}_2(\mathbb{R}^d)$ as a sort of Riemannian manifold as discussed in Subsection 2.3.2. In the second one we discuss three by now classical applications, for which the full power of the abstract theory can be used (i.e. we will have Gradient Flows in the EVI formulation).

Before developing this program, we want to informally discuss a fundamental example.

Let us consider the Entropy functional $E : \mathscr{P}_2(\mathbb{R}^d) \to \mathbb{R} \cup \{+\infty\}$ defined by

$$E(\mu) := \begin{cases} \displaystyle\int \rho \log(\rho) d\mathcal{L}^d, & \text{if } \mu = \rho\mathcal{L}^d, \\ +\infty & \text{otherwise.} \end{cases}$$

We claim that: *the Gradient Flow of the Entropy in* $(\mathscr{P}_2(\mathbb{R}^d), W_2)$ *produces a solution of the Heat equation*. This can be proved rigorously (see Subsection 3.3.2), but for the moment we want to keep the discussion at the heuristic level.

By what discussed in the previous section, we know that the Minimizing Movements scheme produces Gradient Flows. Let us apply the scheme to this setting. Fix an absolutely continuous measure ρ_0 (here we will make no distinction between an absolutely continuous measure and its density), fix $\tau > 0$ and minimize

$$\mu \qquad \mapsto \qquad E(\mu) + \frac{W_2^2(\mu, \rho_0)}{2\tau}. \tag{3.37}$$

It is not hard to see that the minimum is attained at some absolutely continuous measure ρ_τ (actually the minimum is unique, but this has no importance). Our claim will be "proved" if we show that for any $\varphi \in C_c^\infty(\mathbb{R}^d)$ it holds

$$\frac{\int \varphi\rho_\tau - \int \varphi\rho_0}{\tau} = \int \Delta\varphi\, \rho_\tau + o(\tau), \tag{3.38}$$

because this identity tells us that ρ_τ is a first order approximation of the distributional solution of the Heat equation starting from ρ_0 and evaluated at time τ.

To prove (3.38), fix $\varphi \in C_c^\infty(\mathbb{R}^d)$ and perturb ρ_τ in the following way:

$$\rho^\varepsilon := (Id + \varepsilon\nabla\varphi)_\# \rho_\tau.$$

The density of ρ^ε can be explicitly expressed by

$$\rho^\varepsilon(x + \varepsilon\nabla\varphi(x)) = \frac{\rho_\tau(x)}{\det(Id + \varepsilon\nabla^2\varphi(x))}.$$

Observe that it holds

$$\begin{aligned} E(\rho^\varepsilon) &= \int \rho^\varepsilon \log(\rho^\varepsilon) = \int \rho_\tau \log\big(\rho^\varepsilon \circ (Id + \varepsilon\nabla\varphi)\big) = \int \rho_\tau \log\left(\frac{\rho_\tau}{\det(Id + \varepsilon\nabla^2\varphi)}\right) \\ &= E(\rho_\tau) - \int \rho_\tau \log\big(\det(Id + \varepsilon\nabla^2\varphi)\big) = E(\rho_\tau) - \varepsilon \int \rho_\tau \Delta\varphi + o(\varepsilon), \end{aligned} \tag{3.39}$$

where we used the fact that $\det(Id + \varepsilon A) = 1 + \varepsilon \mathrm{tr}(A) + o(\varepsilon)$.

To evaluate the first variation of the distance squared, let T be the optimal transport map from ρ_τ to ρ_0, which exists because of Theorem 1.26, and observe that from $T_\#\rho_\tau = \rho_0$, $(Id + \varepsilon\nabla\varphi)_\#\rho_\tau = \rho^\varepsilon$ and inequality (2.1) we have

$$W_2^2(\rho_0, \rho^\varepsilon) \le \|T - Id - \varepsilon\nabla\varphi\|^2_{L^2(\rho_\tau)},$$

therefore from the fact that equality holds at $\varepsilon = 0$ we get

$$\begin{aligned} W_2^2(\rho_0, \rho^\varepsilon) - W_2^2(\rho_0, \rho_\tau) &\le \|T - Id - \varepsilon\nabla\varphi\|^2_{L^2(\rho_\tau)} - \|T - Id\|^2_{L^2(\rho_\tau)} \\ &= -2\varepsilon \int \langle T - Id, \nabla\varphi\rangle \, \rho_\tau + o(\varepsilon). \end{aligned} \tag{3.40}$$

From the minimality of ρ_τ for the problem (3.37) we know that

$$E(\rho^\varepsilon) + \frac{W_2^2(\rho^\varepsilon, \rho_0)}{2\tau} \ge E(\rho_\tau) + \frac{W_2^2(\rho_\tau, \rho_0)}{2\tau}, \qquad \forall \varepsilon,$$

so that using (3.39) and (3.40), dividing by ε, rearranging the terms and letting $\varepsilon \downarrow 0$ and $\varepsilon \uparrow 0$ we get following Euler-Lagrange equation for ρ_τ:

$$\int \rho_\tau \Delta\varphi + \int \left\langle \frac{T - Id}{\tau}, \nabla\varphi \right\rangle \rho_\tau = 0. \tag{3.41}$$

Now observe that from $T_\#\rho_\tau = \rho_0$ we get

$$\begin{aligned} \frac{\int \varphi\rho_\tau - \int \varphi\rho_0}{\tau} &= -\frac{1}{\tau}\int \big(\varphi(T(x)) - \varphi(x)\big)\rho_\tau(x)dx \\ &= -\frac{1}{\tau}\iint_0^1 \langle \nabla\varphi((1-t)x + tT(x)), T(x) - x\rangle \, dt \, \rho_\tau(x) \, dx \\ &= -\frac{1}{\tau}\int \langle \nabla\varphi(x), T(x) - x\rangle \, \rho_\tau(x) \, dx + \mathrm{Rem}_\tau \\ &\overset{(3.41)}{=} \int \Delta\varphi \, \rho_\tau + \mathrm{Rem}_\tau, \end{aligned}$$

where the remainder term Rem_τ is bounded by

$$|\mathrm{Rem}_\tau| \le \frac{\mathrm{Lip}(\nabla\varphi)}{\tau} \iint_0^1 t|T(x) - x|^2 dt \, \rho_\tau(x) \, dx = \frac{\mathrm{Lip}(\nabla\varphi)}{2\tau} W_2^2(\rho_0, \rho_\tau).$$

Since, heuristically speaking, $W_2(\rho_0, \rho_\tau)$ has the same magnitude of τ, we have $\mathrm{Rem}_\tau = o(\tau)$ and the "proof" is complete.

3.3.1 Elements of subdifferential calculus in $(\mathscr{P}_2(\mathbb{R}^d), W_2)$

Recall that we introduced a weak Riemannian structure on the space $(\mathscr{P}_2(M), W_2)$ in Subsection 2.3.2. Among others, this weak Riemannian structure of $(\mathscr{P}_2(M), W_2)$ allows the development of a *subdifferential calculus for geodesically convex functionals*, in the same spirit (and with many formal similarities) of the usual subdifferential calculus for convex functionals on an Hilbert space.

To keep the notation and the discussion simpler, we are going to define the subdifferential of a geodesically convex functional only for the case $\mathscr{P}_2(\mathbb{R}^d)$ and for regular measures (Definition 1.25), but everything can be done also on manifolds (or Hilbert spaces) and for general $\mu \in \mathscr{P}_2(M)$.

Recall that for a λ-convex functional F on an Hilbert space H, the subdifferential $\partial^- F(x)$ at a point x is the set of vectors $v \in H$ such that

$$F(x) + \langle v, y - x\rangle + \frac{\lambda}{2}|x - y|^2 \leq F(y), \qquad \forall y \in H.$$

Definition 3.26 (Subdifferential in $(\mathscr{P}_2(\mathbb{R}^d), W_2)$) *Let $E : \mathscr{P}_2(\mathbb{R}^d) \to \mathbb{R} \cup \{+\infty\}$ be a λ-geodesically convex and lower semicontinuous functional, and $\mu \in \mathscr{P}_2(\mathbb{R}^d)$ be a regular measure such that $E(\mu) < \infty$. The set $\partial^W E(\mu) \subset \mathrm{Tan}_\mu(\mathscr{P}_2(\mathbb{R}^d))$ is the set of vector fields $v \in L^2(\mu, \mathbb{R}^d)$ such that*

$$E(\mu) + \int \langle T_\mu^\nu - Id, v\rangle \, d\mu + \frac{\lambda}{2} W_2^2(\mu, \nu) \leq E(\nu), \qquad \forall \nu \in \mathscr{P}_2(\mathbb{R}^d),$$

where here and in the following T_μ^ν will denote the optimal transport map from the regular measure μ to ν (whose existence and uniqueness is guaranteed by Theorem 1.26).

Observe that the subdifferential of a λ-geodesically convex functional E has the following monotonicity property (which closely resembles the analogous valid for λ-convex functionals on an Hilbert space):

$$\int \langle v, T_\mu^\nu - Id\rangle \, d\mu + \int \langle w, T_\nu^\mu - Id\rangle \, d\nu \leq -\lambda W_2^2(\mu, \nu), \tag{3.42}$$

for every couple of regular measures μ, ν in the domain of E, and $v \in \partial^W E(\mu)$, $w \in \partial^W E(\nu)$. To prove (3.42) just observe that from the definition of subdifferential we have

$$\begin{aligned} E(\mu) + \int \langle T_\mu^\nu - Id, v\rangle \, d\mu + \frac{\lambda}{2} W_2^2(\mu, \nu) &\leq E(\nu), \\ E(\nu) + \int \langle T_\nu^\mu - Id, w\rangle \, d\nu + \frac{\lambda}{2} W_2^2(\mu, \nu) &\leq E(\mu), \end{aligned}$$

and add up these inequalities.

The definition of subdifferential leads naturally to the definition of Gradient Flow: it is sufficient to transpose the definition given with the system (3.2).

Definition 3.27 (Subdifferential formulation of Gradient Flow) *Let E be a λ-geodesically convex functional on $\mathscr{P}_2(\mathbb{R}^d)$ and $\mu \in \mathscr{P}_2(\mathbb{R}^d)$. Then (μ_t) is a Gradient Flow for E starting from μ provided it is a locally absolutely continuous curve, $\mu_t \to \mu$ as $t \to 0$ w.r.t. the distance W_2, μ_t is regular for $t > 0$ and it holds*

$$-v_t \in \partial^W E(\mu_t), \qquad a.e.\ t,$$

where (v_t) is the vector field uniquely identified by the curve (μ_t) via

$$\begin{aligned} &\frac{d}{dt}\mu_t + \nabla \cdot (v_t \mu_t) = 0, \\ &v_t \in \mathrm{Tan}_{\mu_t}(\mathscr{P}_2(\mathbb{R}^d)) \qquad a.e.\ t, \end{aligned}$$

(recall Theorem 2.29 and Definition 2.31).

Thus we have a total of 4 different formulations of Gradient Flows of λ-geodesically convex functionals on $\mathscr{P}_2(\mathbb{R}^d)$ based respectively on the Energy Dissipation Inequality, the Energy Dissipation Equality, the Evolution Variational Inequality and the notion of subdifferential.

The important point is that these 4 formulations are *equivalent* for $\lambda-$geodesically convex functionals:

Proposition 3.28 (Equivalence of the various formulation of GF in the Wasserstein space) *Let E be a λ-geodesically convex functional on $\mathscr{P}_2(\mathbb{R}^d)$ and (μ_t) a curve made of regular measures. Then for (μ_t) the 4 definitions of Gradient Flow for E (EDI, EDE, EVI and the Subdifferential one) are equivalent.*

Sketch of the Proof
We prove only that the EVI formulation is equivalent to the Subdifferential one. Recall that by Proposition 2.34 we know that

$$\frac{1}{2}\frac{d}{dt}W_2^2(\mu_t,\nu) = -\int \langle v_t, T_{\mu_t}^{\nu} - Id\rangle\, d\mu_t, \qquad a.e.t$$

where $T_{\mu_t}^{\nu}$ is the optimal transport map from μ_t to ν. Then we have

$$\begin{gathered}
-v_t \in \partial^W E(\mu_t), \qquad a.e.\ t,\\
\Updownarrow\\
E(\mu_t) + \int \langle -v_t, T_{\mu_t}^{\nu} - Id\rangle\, d\mu_t + \frac{\lambda}{2}W_2^2(\mu_t,\nu) \le E(\nu), \qquad \forall \nu \in \mathscr{P}_2(\mathbb{R}^d),\ a.e.\ t\\
\Updownarrow\\
E(\mu_t) + \frac{1}{2}\frac{d}{dt}W_2^2(\mu_t,\nu) + \frac{\lambda}{2}W_2^2(\mu_t,\nu) \le E(\nu), \qquad \forall \nu \in \mathscr{P}_2(\mathbb{R}^d),\ a.e.\ t.
\end{gathered}$$

□

3.3.2 Three classical functionals

We now pass to the analysis of 3 by now classical examples of Gradient Flows in the Wasserstein space. Recall that in terms of strength, the best theory to use is the one of Subsection 3.2.4, because the compatibility in Energy and distance ensures strong properties both at the level of discrete solutions and for the limit curve obtained. Once we will have a Gradient Flow, the Subdifferential formulation will let us understand which is the PDE associated to it.

Let us recall (Example 2.21) that the space $(\mathscr{P}_2(\mathbb{R}^d), W_2)$ is *not* Non Positively Curved in the sense of Alexandrov, this means that if we want to check whether a given functional is compatible with the distance or not, we cannot use geodesics to interpolate between points (because we would violate the second inequality in (3.29)). A priori the choice of the interpolating curves may depend on the functional, but actually in what comes next we will always use the ones defined by:

Definition 3.29 (Interpolating curves) *Let μ, ν_0, $\nu_1 \in \mathscr{P}_2(\mathbb{R}^d)$ and assume that μ is regular (Definition 1.25). The interpolating curve (ν_t) from ν_0 to ν_1 with base μ is defined as*

$$\nu_t := ((1-t)T_0 + tT_1)_{\#}\mu,$$

where T_0 and T_1 are the optimal transport maps from μ to ν_0 and ν_1 respectively. Observe that if $\mu = \nu_0$, the interpolating curve reduces to the geodesic connecting it to ν_1.

Strictly speaking, in order to apply the theory of Section 3.2.4 we should define interpolating curves having as base any measure $\mu \in \mathscr{P}_2(\mathbb{R}^d)$, and not just regular ones. This is actually possible, and the foregoing discussion can be applied to the more general definition, but we prefer to avoid technicalities, and just focus on the main concepts.

For an interpolating curve as in the definition it holds:

$$W_2^2(\mu, \nu_t) \leq (1-t)W_2^2(\mu, \nu_0) + tW_2^2(\mu, \nu_1) - t(1-t)W_2^2(\nu_0, \nu_1). \tag{3.43}$$

Indeed the map $(1-t)T_0 + tT_1$ is optimal from μ to ν_t (because we know that T_0 and T_1 are the gradients of convex functions φ_0, φ_1 respectively, thus $(1-t)T_0 + tT_1$ is the gradient of the convex function $(1-t)\varphi_0 + t\varphi_1$, and thus is optimal), and we know by inequality (2.1) that $W_2^2(\nu_0, \nu_1) \leq \|T_0 - T_1\|^2_{L^2(\mu)}$, thus it holds

$$\begin{aligned} W_2^2(\mu, \nu_t) &= \|(1-t)T_0 + tT_1\|^2_{L^2(\mu)} \\ &= (1-t)\|T_0 - Id\|^2_{L^2(\mu)} + t\|T_1 - Id\|^2_{L^2(\mu)} - t(1-t)\|T_0 - T_1\|^2_{L^2(\mu)} \\ &\leq (1-t)W_2^2(\mu, \nu_0) + tW_2^2(\mu, \nu_1) - t(1-t)W_2^2(\nu_0, \nu_1). \end{aligned}$$

We now pass to the description of the three functionals we want to study.

Definition 3.30 (Potential energy) *Let $V : \mathbb{R}^d \to \mathbb{R} \cup \{+\infty\}$ be lower semicontinuous and bounded from below. The potential energy functional $\mathcal{V} : \mathscr{P}_2(\mathbb{R}^d) \to \mathbb{R} \cup \{+\infty\}$ associated to V is defined by*

$$\mathcal{V}(\mu) := \int V d\mu.$$

Definition 3.31 (Interaction energy) *Let $W : \mathbb{R}^d \to \mathbb{R} \cup \{+\infty\}$ be lower semicontinuous, even and bounded from below. The interaction energy functional $\mathcal{W} : \mathscr{P}_2(\mathbb{R}^d) \to \mathbb{R} \cup \{+\infty\}$ associated to W is defined by*

$$\mathcal{W}(\mu) := \frac{1}{2} \int W(x_1 - x_2) d\mu \times \mu(x_1, x_2).$$

Observe that the definition makes sense also for not even functions W; however, replacing if necessary the function $W(x)$ with $(W(x) + W(-x))/2$ we get an even function leaving the value of the functional unchanged.

Definition 3.32 (Internal energy) *Let $u : [0, +\infty) \to \mathbb{R} \cup \{+\infty\}$ be a convex function bounded from below such that $u(0) = 0$ and*

$$\varliminf_{z \to 0} \frac{u(z)}{z^\alpha} > -\infty, \qquad \text{for some } \alpha > \frac{d}{d+2}, \tag{3.44}$$

let $u'(\infty) := \lim_{z \to \infty} u(z)/z$. The internal energy functional $\mathcal{E}$ associated to u is

$$\mathcal{E}(\mu) := \int u(\rho)\mathcal{L}^d + u'(\infty)\mu^s(\mathbb{R}^d),$$

where $\mu = \rho\mathcal{L}^d + \mu^s$ is the decomposition of μ in absolutely continuous and singular parts w.r.t. the Lebesgue measure.

Condition (3.44) ensures that the negative part of $u(\rho)$ is integrable for $\mu \in \mathscr{P}_2(\mathbb{R}^d)$, so that $\mathcal{E}$ is well defined (possibly $+\infty$). Indeed from (3.44) we have $u^-(z) \leq az + bz^\alpha$ for some $\alpha < 1$ satisfying $2\alpha/(1-\alpha) > d$, and it holds

$$\begin{aligned}
\int \rho^\alpha(x) d\mathcal{L}^d(x) &= \int \rho^\alpha(x)(1+|x|)^{2\alpha}(1+|x|)^{-2\alpha} d\mathcal{L}^d(x) \\
&\leq \left(\int \rho(x)(1+|x|)^2 d\mathcal{L}^d(x)\right)^\alpha \left(\int (1+|x|)^{\frac{-2\alpha}{1-\alpha}} \mathcal{L}^d(x)\right)^{1-\alpha} < \infty.
\end{aligned}$$

Under appropriate assumptions on V, W and e the above defined functionals are compatible with the distance W_2. As said before we will use as interpolating curves those given in Definition 3.29.

Proposition 3.33 *Let $\lambda \geq 0$. The following holds.*

i) The functional $\mathcal{V}$ is λ-convex along interpolating curves in $(\mathscr{P}_2(\mathbb{R}^d), W_2)$ if and only if V is λ-convex.

ii) The functional $\mathcal{W}$ is λ-convex along interpolating curves $(\mathscr{P}_2(\mathbb{R}^d), W_2)$ if W is λ-convex.

iii) The functional $\mathcal{E}$ is convex along interpolating curves $(\mathscr{P}_2(\mathbb{R}^d), W_2)$ provided u satisfies

$$z \quad \mapsto \quad z^d u(z^{-d}) \qquad \text{is convex and non increasing on } (0, +\infty). \tag{3.45}$$

Proof Since the second inequality in (3.29) is satisfied by the interpolating curves that we are considering (inequality (3.43)) we need only to check the convexity of the functionals.

Let (ν_t) be an interpolating curve with base the regular measure μ, and T_0, T_1 the optimal transport maps from μ to ν_0 and ν_1 respectively.

The *only if* part of (i) follows simply considering interpolation of deltas. For the *if*, observe that[5]

$$\begin{aligned}
\mathcal{V}(\nu_t) &= \int V(x) d\nu_t(x) = \int V\big((1-t)T_0(x) + tT_1(x)\big) d\mu(x) \\
&\leq (1-t)\int V(T_0(x))d\mu(x) + t\int V(T_1(x))d\mu(x) - \frac{\lambda}{2}t(1-t)\int |T_0(x) - T_1(x)|^2 d\mu(x) \\
&\leq (1-t)\mathcal{V}(\nu_0) + t\mathcal{V}(\nu_1) - \frac{\lambda}{2}t(1-t)W_2^2(\nu_0,\nu_1).
\end{aligned} \tag{3.46}$$

For (ii) we start claiming that $W_2^2(\mu\times\mu, \nu\times\nu) = 2W_2^2(\mu,\nu)$ for any $\mu,\nu \in \mathscr{P}_2(\mathbb{R}^d)$. To prove this, it is enough to check that if $\boldsymbol{\gamma} \in \mathit{Opt}(\mu,\nu)$ then $\tilde{\boldsymbol{\gamma}} := (\pi^1,\pi^1,\pi^2,\pi^2)_\#\boldsymbol{\gamma} \in \mathit{Opt}(\mu\times\mu, \nu\times\nu)$. To see this, let $\varphi : \mathbb{R}^d \to \mathbb{R}\cup\{+\infty\}$ be a convex function such that $\mathrm{supp}(\boldsymbol{\gamma}) \subset \partial^-\varphi$ and define the convex function $\tilde{\varphi}$ on $\mathbb{R}^{2d}$ by $\tilde{\varphi}(x,y) = \varphi(x)+\varphi(y)$. It is immediate to verify that $\mathrm{supp}(\tilde{\boldsymbol{\gamma}}) \subset \partial^-\tilde{\varphi}$, so that $\tilde{\boldsymbol{\gamma}}$ is optimal as well. This argument also shows that if (ν_t) is an interpolating curve with base μ, then $t \mapsto \nu_t \times \nu_t$ is an interpolating curve from $\nu_0\times\nu_0$ to $\nu_1\times\nu_1$ with base $\mu\times\mu$. Also, $(x_1,x_2) \mapsto W(x_1 - x_2)$ is λ-convex if W is. The conclusion now follows from case (i).

We pass to (iii). We will make the simplifying assumption that $\mu \ll \mathcal{L}^d$ and that T_0 and T_1 are smooth and satisfy $\det(\nabla T_0)(x) \neq 0$, $\det(\nabla T_1)(x) \neq 0$ for every $x \in \mathrm{supp}(\mu)$ (up to an approximation argument, it is possible to reduce to this case, we omit the details). Then, writing $\mu = \rho\mathcal{L}^d$, from the change of variable formula we get that $\nu_t \ll \mathcal{L}^d$ and for its density $\tilde{\rho}_t$ it holds

$$\tilde{\rho}_t(T_t(x)) = \frac{\rho(x)}{\det(\nabla T_t(x))},$$

[5]the assumption $\lambda \geq 0$ is necessary to have the last inequality in (3.46). If $\lambda < 0$, $\lambda-$convexity of $\mathcal{V}$ along interpolating curves is not anymore true, so that we cannot apply directly the results of Subsection 3.2.4. Yet, adapting the arguments, it possible to show that all the results which we will present hereafter are true for general $\lambda \in \mathbb{R}$.

where we wrote T_t for $(1-t)T_0 + tT_1$. Thus

$$\mathcal{E}(\nu_t) = \int u(\tilde{\rho}_t(y))d\mathcal{L}^d(y) = \int u\left(\frac{\rho(x)}{\det(\nabla T_t)(x)}\right)\det(\nabla T_t)(x)d\mathcal{L}^d(x).$$

Therefore the proof will be complete if we show that $A \mapsto u(\frac{\rho(x)}{\det(A)})\det(A)$ is convex on the set of positively defined symmetric matrices for any $x \in \mathrm{supp}(\mu)$. Observe that this map is the composition of the convex and non increasing map $z \mapsto z^d u(\rho(x)/z^d)$ with the map $A \mapsto (\det(A))^{1/d}$. Thus to conclude it is sufficient to show that $A \mapsto (\det(A))^{1/d}$ is concave. To this aim, pick two symmetric and positive definite matrices A_0 and A_1, notice that

$$\big(\det((1-t)A_0 + tA_1)\big)^{1/d} = \big(\det(A_0)\det(Id + tB)\big)^{1/d},$$

where $B = \sqrt{A_0}(A_1 - A_0)\sqrt{A_0}$ and conclude by

$$\begin{aligned}
\frac{d}{dt}\det(Id+tB)^{1/d} &= \frac{1}{d}\big(\det(Id+tB)\big)^{1/d}\mathrm{tr}\big(B\,(Id+tB)^{-1}\big),\\
\frac{d^2}{dt^2}\det(Id+tB)^{1/d} &= \frac{1}{d^2}\mathrm{tr}^2\big(B\,(Id+tB)^{-1}\big) - \frac{1}{d}\mathrm{tr}\big(\big(B\,(Id+tB)^{-1}\big)^2\big) \le 0
\end{aligned}$$

where in the last step we used the inequality $\mathrm{tr}^2(C) \le d\,\mathrm{tr}(C^2)$ for $C = B\,(Id+tB)^{-1}$. □

Important examples of functions u satisfying (3.44) and (3.45) are:

$$\begin{aligned}
u(z) &= \frac{z^\alpha - z}{\alpha - 1}, \qquad \alpha \ge 1 - \frac{1}{d},\ \alpha \neq 1\\
u(x) &= z\log(z).
\end{aligned} \tag{3.47}$$

Remark 3.34 (A dimension free condition on u) We saw that a sufficient condition on u to ensure that $\mathcal{E}$ is convex along interpolating curves is the fact that the map $z \mapsto z^d u(z^{-d})$ is convex and non increasing, so the dimension d of the ambient space plays a role in the condition. The fact that the map is non increasing follows by the convexity of u together with $u(0) = 0$, while by simple computations we see that its convexity is equivalent to

$$z^{-1}u(z) - u'(z) + zu''(z) \ge -\frac{1}{d-1}zu''(z). \tag{3.48}$$

Notice that the higher d is, the stricter the condition becomes. For applications in infinite dimensional spaces, it is desirable to have a condition on u ensuring the convexity of $\mathcal{E}$ in which the dimension does not enter. As inequality (3.48) shows, the weakest such condition for which $\mathcal{E}$ is convex in any dimension is:

$$z^{-1}u(z) - u'(z) + zu''(z) \ge 0,$$

and some computations show that this is in turn equivalent to the convexity of the map

$$z \quad \mapsto \quad e^z u(e^{-z}).$$

A key example of map satisfying this condition is $z \mapsto z\log(z)$. ■

Therefore we have the following existence and uniqueness result:

Theorem 3.35 *Let $\lambda \geq 0$ and $\mathcal{F}$ be either $\mathcal{V}$, $\mathcal{W}$, $\mathcal{E}$ (or a linear combination of them with positive coefficients) and λ-convex along interpolating curves. Then for every $\overline{\mu} \in \mathscr{P}_2(\mathbb{R}^d)$ there exists a unique Gradient Flow (μ_t) for $\mathcal{F}$ starting from $\overline{\mu}$ in the EVI formulation. The curve (μ_t) satisfies: is locally absolutely continuous on $(0, +\infty)$, $\mu_t \to \overline{\mu}$ as $t \to 0$ and, if μ_t is regular for every $t \geq 0$, it holds*

$$-v_t \in \partial^W F(\mu_t), \qquad a.e.\ t \in (0, +\infty), \tag{3.49}$$

where (v_t) is the velocity vector field associated to (μ_t) characterized by

$$\frac{d}{dt}\mu_t + \nabla \cdot (v_t \mu_t) = 0,$$
$$v_t \in \operatorname{Tan}_{\mu_t}(\mathscr{P}_2(\mathbb{R}^d)) \qquad a.e.\ t.$$

Proof Use the existence Theorem 3.25 and the equivalence of the EVI formulation of Gradient Flow and the Subdifferential one provided by Proposition 3.28. □

It remains to understand which kind of equation is satisfied by the Gradient Flow (μ_t). By equation (3.49), this corresponds to identify the subdifferentials of $\mathcal{V}$, $\mathcal{W}$, $\mathcal{E}$ at a generic $\mu \in \mathscr{P}_2(\mathbb{R}^d)$. This is the content of the next three propositions. For simplicity, we state and prove them only under some - unneeded - smoothness assumptions. The underlying idea of all the calculations we are going to do is the following equivalence:

$$v \in \partial^W \mathcal{F}(\mu) \qquad \overset{\approx}{\Leftrightarrow} \qquad \lim_{\varepsilon \to 0} \frac{\mathcal{F}((Id + \varepsilon\nabla\varphi)_\# \mu) - \mathcal{F}(\mu)}{\varepsilon} = \int \langle v, \nabla\varphi \rangle\,,\ \forall \varphi \in C_c^\infty(\mathbb{R}^d), \tag{3.50}$$

valid for any λ-geodesically convex functional, where we wrote $\overset{\approx}{\Leftrightarrow}$ to intend that this equivalence holds only when everything is smooth. To understand why (3.50) holds, start assuming that $v \in \partial^W F(\mu)$, fix $\varphi \in C_c^\infty(\mathbb{R}^d)$ and recall that for ε sufficiently small the map $Id + \varepsilon\nabla\varphi$ is optimal (Remark 1.22). Thus by definition of subdifferential we have

$$\mathcal{F}(\mu) + \varepsilon \int \langle v, \nabla\varphi \rangle\, d\mu + \varepsilon^2 \frac{\lambda}{2} \|\nabla\varphi\|^2_{L^2(\mu)} \leq \mathcal{F}((Id + \varepsilon\nabla\varphi)_\# \mu).$$

Subtracting $\mathcal{F}(\mu)$ on both sides, dividing by $\varepsilon > 0$ and $\varepsilon < 0$ and letting $\varepsilon \to 0$ we get the implication $\Rightarrow$. To "prove" the converse one, pick $\nu \in \mathscr{P}_2(\mathbb{R}^d)$, let T be the optimal transport map from μ to ν and recall that T is the gradient of a convex function ϕ. Assume that ϕ is smooth and define $\varphi(x) := \phi(x) - |x|^2/2$. The geodesic (μ_t) from μ to ν can then be written as

$$\mu_t = \big((1-t)Id + tT\big)_\# \mu = \big((1-t)Id + t\nabla\phi\big)_\# \mu = \big(Id + t\nabla\varphi\big)_\# \mu.$$

From the λ-convexity hypothesis we know that

$$\mathcal{F}(\nu) \geq \mathcal{F}(\mu) + \frac{d}{dt}|_{t=0} \mathcal{F}(\mu_t) + \frac{\lambda}{2} W_2^2(\mu, \nu),$$

therefore, since we know that $\frac{d}{dt}|_{t=0}\mathcal{F}(\mu_t) = \int \langle v, \nabla\varphi \rangle\, d\mu$, from the arbitrariness of ν we deduce $v \in \partial^W \mathcal{F}(\mu)$.

Proposition 3.36 (Subdifferential of $\mathcal{V}$) *Let $V : \mathbb{R}^d \to \mathbb{R}$ be λ-convex and C^1, let $\mathcal{V}$ be as in Definition 3.30 and let $\mu \in \mathscr{P}_2(\mathbb{R}^d)$ be regular and satisfying $\mathcal{V}(\mu) < \infty$. Then $\partial^W \mathcal{V}(\mu)$ is non empty if and only if $\nabla V \in L^2(\mu)$, and in this case ∇V is the only element in the subdifferential of $\mathcal{V}$ at μ.*

Therefore, if (μ_t) is a Gradient Flow of $\mathcal{V}$ made of regular measures, it solves

$$\frac{d}{dt}\mu_t = \nabla \cdot (\nabla V \mu_t),$$

in the sense of distributions in $\mathbb{R}^d \times (0, +\infty)$.

Sketch of the Proof Fix $\varphi \in C_c^\infty(\mathbb{R}^d)$ and observe that

$$\lim_{\varepsilon \to 0} \frac{\mathcal{V}((Id + \varepsilon \nabla \varphi)_\# \mu) - \mathcal{V}(\mu)}{\varepsilon} = \lim_{\varepsilon \to 0} \int \frac{V \circ (Id + \varepsilon \nabla \varphi) - V}{\varepsilon} d\mu = \int \langle \nabla V, \nabla \varphi \rangle \, d\mu.$$

Conclude using the equivalence (3.50). □

Proposition 3.37 (Subdifferential of $\mathcal{W}$) *Let $W : \mathbb{R}^d \to \mathbb{R}$ be λ-convex, even and C^1, let $\mathcal{W}$ be defined by 3.31 and μ be regular and satisfying $\mathcal{W}(\mu) < \infty$. Then $\partial^W \mathcal{W}(\mu) \neq \emptyset$ if and only if $(\nabla W) * \mu$ belongs to $L^2(\mu)$ and in this case $(\nabla W) * \mu$ is the only element in the subdifferential of $\mathcal{W}$ at μ.*

Therefore, if (μ_t) is a Gradient Flow of $\mathcal{W}$ made of regular measures, it solves the non local evolution equation

$$\frac{d}{dt}\mu_t = \nabla \cdot ((\nabla W * \mu_t)\mu_t),$$

in the sense of distributions in $\mathbb{R}^d \times (0, +\infty)$.

Sketch of the Proof Fix $\varphi \in C_c^\infty(\mathbb{R}^d)$, let $\mu^\varepsilon := (Id + \varepsilon \nabla \varphi)_\# \mu$ and observe that

$$\begin{aligned}
\mathcal{W}(\mu^\varepsilon) &= \frac{1}{2}\int W(x-y) d\mu^\varepsilon(x) d\mu^\varepsilon(y) = \frac{1}{2}\int W(x - y + \varepsilon(\nabla\varphi(x) - \nabla\varphi(y))) d\mu(x) d\mu(y) \\
&= \frac{1}{2}\int W(x-y) d\mu(x) d\mu(y) + \frac{\varepsilon}{2}\int \langle \nabla W(x-y), \nabla\varphi(x) - \nabla\varphi(y) \rangle \, d\mu(x) d\mu(y) + o(\varepsilon).
\end{aligned}$$

Now observe that

$$\begin{aligned}
\int \langle \nabla W(x-y), \nabla\varphi(x) \rangle \, d\mu(x) d\mu(y) &= \int \left\langle \int \nabla W(x-y) d\mu(y), \nabla\varphi(x) \right\rangle d\mu(x) \\
&= \int \langle \nabla W * \mu(x), \nabla\varphi(x) \rangle \, d\mu(x),
\end{aligned}$$

and, similarly,

$$\begin{aligned}
\int \langle \nabla W(x-y), -\nabla\varphi(y) \rangle \, d\mu(x) d\mu(y) &= \int \langle \nabla W * \mu(y), \nabla\varphi(y) \rangle \, d\mu(y) \\
&= \int \langle \nabla W * \mu(x), \nabla\varphi(x) \rangle \, d\mu(x).
\end{aligned}$$

Thus the conclusion follows by applying the equivalence (3.50). □

Proposition 3.38 (Subdifferential of $\mathcal{E}$) *Let $u : [0, +\infty) \to \mathbb{R}$ be convex, C^2 on $(0, +\infty)$, bounded from below and satisfying conditions* (3.44) *and* (3.45)*. Let $\mu = \rho \mathcal{L}^d \in \mathscr{P}_2(\mathbb{R}^d)$ be an absolutely continuous measure with smooth density. Then $\nabla(u'(\rho))$ is the unique element in $\partial^W \mathcal{E}(\mu)$.*

Therefore, if (μ_t) is a Gradient Flow for $\mathcal{E}$ and μ_t is absolutely continuous with smooth density ρ_t for every $t > 0$, then $t \mapsto \rho_t$ solves the equation

$$\frac{d}{dt}\rho_t = \nabla \cdot (\rho_t \nabla (u'(\rho_t))).$$

Note: this statement is not perfectly accurate, because we are neglecting the integrability issues. Indeed a priori we don't know that $\nabla(u'(\rho))$ belongs to $L^2(\mu)$.

Sketch of the Proof Fix $\varphi \in C_c^\infty(\mathbb{R}^d)$ and define $\mu^\varepsilon := (Id + \varepsilon\nabla\varphi)_\#\mu$. For ε sufficiently small, μ^ε is absolutely continuous and its density ρ^ε satisfies - by the change of variable formula - the identity

$$\rho^\varepsilon(x + \varepsilon\nabla\varphi(x)) = \frac{\rho(x)}{\det(Id + \varepsilon\nabla^2\varphi(x))}.$$

Using the fact that $\frac{d}{d\varepsilon}|_{\varepsilon=0}(\det(Id + \varepsilon\nabla^2\varphi(x))) = \Delta\varphi(x)$ we have

$$\begin{aligned}\frac{d}{d\varepsilon}|_{\varepsilon=0}\mathcal{E}(\mu^\varepsilon) &= \frac{d}{d\varepsilon}|_{\varepsilon=0}\int u(\rho^\varepsilon(y))dy = \frac{d}{d\varepsilon}|_{\varepsilon=0}\int u\left(\frac{\rho(x)}{\det(Id + \varepsilon\nabla^2\varphi(x))}\right)\det(Id + \varepsilon\nabla^2\varphi(x))dx \\ &= \int -\rho u'(\rho)\Delta\varphi + u(\rho)\Delta\varphi = \int \langle\nabla(\rho u'(\rho) - u(\rho)), \nabla\varphi\rangle = \int \langle\nabla(u'(\rho)), \nabla\varphi\rangle\,\rho,\end{aligned}$$

and the conclusion follows by the equivalence (3.50). □

As an example, let $u(z) := z\log(x)$, and let V be a λ-convex smooth function on $\mathbb{R}^d$. Since $u'(z) = \log(z) + 1$, we have $\rho\nabla(u'(\rho)) = \Delta\rho$, thus a gradient flow (ρ_t) of $\mathcal{F} = \mathcal{E} + \mathcal{V}$ solves the Fokker-Plank equation

$$\frac{d}{dt}\rho_t = \Delta\rho_t + \nabla\cdot(\nabla V\rho_t).$$

Also, the contraction property (3.31) in Theorem 3.25 gives that for two gradient flows (ρ_t), $(\tilde\rho_t)$ it holds the contractivity estimate

$$W_2(\rho_t, \tilde\rho_t) \le e^{-\lambda t}W_2(\rho_0, \tilde\rho_0).$$

3.4 Bibliographical notes

The content of Section 3.2 is taken from the first part of [6] (we refer to this book for a detailed bibliographical references on the topic of gradient flows in metric spaces), with the only exception of Proposition 3.6, whose proof has been communicated to us by Savaré (see also [72], [73]).

The study of geodesically convex functionals in $(\mathscr{P}_2(\mathbb{R}^d), W_2)$ has been introduced by R. McCann in [63], who also proved that conditions (3.44) and (3.45) were sufficient to deduce the geodesic convexity (called by him displacement convexity) of the internal energy functional.

The study of gradient flows in the Wasserstein space began in the seminal paper by R. Jordan, D. Kinderlehrer and F. Otto [47], where it was proved that the minimizing movements procedure for the functional

$$\rho\mathcal{L}^d \qquad \mapsto \qquad \int \rho\log\rho + V\rho d\mathcal{L}^d,$$

on the space $(\mathscr{P}_2(\mathbb{R}^d), W_2)$, produce solutions of the Fokker-Planck equation. Later, F. Otto in [67] showed that the same discretization applied to

$$\rho\mathcal{L}^d \qquad \mapsto \qquad \frac{1}{\alpha - 1}\int \rho^\alpha d\mathcal{L}^d,$$

(with the usual meaning for measures with a singular part) produce solutions of the porous medium equation. The impact of Otto's work on the community of optimal transport has been huge: not only he was able to provide concrete consequences (in terms of new estimates for the rate of convergence of solutions of the porous medium equation) out of optimal transport theory, but he also clearly described what is now called the 'weak Riemannian structure' of $(\mathscr{P}_2(\mathbb{R}^d), W_2)$ (see also Chapter 6 and Subsection 2.3.2).

Otto's intuitions have been studied and extended by many authors. The rigorous description of many of the objects introduced by Otto, as well as a general discussion about gradient flows of λ-geodesically convex functionals on $(\mathscr{P}_2(\mathbb{R}^d), W_2)$ has been done in the second part of [6] (the discussion made here is taken from this latter reference).

4 Geometric and functional inequalities

In this short Chapter we show how techniques coming from optimal transport can lead to simple proofs of some important geometric and functional inequalities. None of the results proven here are new, in the sense that they all were well known before the proofs coming from optimal transport appeared. Still, it is interesting to observe how the tools described in the previous sections allow to produce proofs which are occasionally simpler and in any case providing new informations when compared to the 'standard' ones.

4.1 Brunn-Minkowski inequality

Recall that the Brunn-Minkowski inequality in $\mathbb{R}^d$ is:

$$\left(\mathscr{L}^d\left(\frac{A+B}{2}\right)\right)^{1/d} \geq \frac{1}{2}\left(\left(\mathscr{L}^d(A)\right)^{1/d} + \left(\mathscr{L}^d(B)\right)^{1/d}\right),$$

and is valid for any couple of compact sets $A,\ B \subset \mathbb{R}^d$.

To prove it, let $A, B \subset \mathbb{R}^d$ be compact sets and notice that without loss of generality we can assume that $\mathscr{L}^d(A), \mathscr{L}^d(B) > 0$. Define

$$\mu_0 := \frac{1}{\mathscr{L}^d(A)}\mathscr{L}^d|_A \qquad \mu_1 := \frac{1}{\mathscr{L}^d(B)}\mathscr{L}^d|_B,$$

and let (μ_t) be the unique geodesic in $(\mathscr{P}_2(\mathbb{R}^d), W_2)$ connecting them.

Recall from (3.47) that for $u(z) = -d(z^{1-1/d} - z)$ the functional $\mathcal{E}(\rho) := \int u(\rho)d\mathcal{L}^d$ is geodesically convex in $(\mathscr{P}_2(\mathbb{R}^d), W_2)$. Also, simple calculations show that $\mathcal{E}(\mu_0) = -d(\mathscr{L}^d(A)^{1/d} - 1)$, $\mathcal{E}(\mu_1) = -d(\mathscr{L}^d(B)^{1/d} - 1)$. Hence we have

$$\mathcal{E}(\mu_{1/2}) \leq -\frac{d}{2}\Big(\left(\mathscr{L}^d(A)\right)^{1/d} + \left(\mathscr{L}^d(B)\right)^{1/d}\Big) + d.$$

Now notice that Theorem 2.10 (see also Remark 2.13) ensures that $\mu_{1/2}$ is concentrated on $\frac{A+B}{2}$, thus letting $\tilde{\mu}_{1/2} := (\mathscr{L}^d((A+B)/2))^{-1}\mathscr{L}^d|_{(A+B)/2}$ and applying Jensen's inequality to the convex function u we get

$$\mathcal{E}(\mu_{1/2}) \geq \mathcal{E}(\tilde{\mu}_{1/2}) = -d\left(\mathscr{L}^d\Big(\frac{A+B}{2}\Big)^{1/d} - 1\right),$$

which concludes the proof.

4.2 Isoperimetric inequality

On $\mathbb{R}^d$ the isoperimetric inequality can be written as

$$\mathscr{L}^d(E)^{1-\frac{1}{d}} \le \frac{P(E)}{d\mathscr{L}^d(B)^{\frac{1}{d}}},$$

where E is an arbitrary open set, $P(E)$ its perimeter and B the unitary ball.

We will prove this inequality via Brenier's theorem 1.26, neglecting all the smoothness issues. Let

$$\mu := \frac{1}{\mathscr{L}^d(E)}\mathscr{L}^d|_E, \qquad \nu := \frac{1}{\mathscr{L}^d(B)}\mathscr{L}^d|_B,$$

and $T : E \to B$ be the optimal transport map (w.r.t. the cost given by the distance squared). The change of variable formula gives

$$\frac{1}{\mathscr{L}^d(E)} = \det(\nabla T(x))\frac{1}{\mathscr{L}^d(B)}, \qquad \forall x \in E.$$

Since we know that T is the gradient of a convex function, we have that $\nabla T(x)$ is a symmetric matrix with non negative eigenvalues for every $x \in E$. Hence the arithmetic-geometric mean inequality ensures that

$$(\det \nabla T(x))^{1/d} \le \frac{\nabla \cdot T(x)}{d}, \qquad \forall x \in E.$$

Coupling the last two equations we get

$$\frac{1}{\mathscr{L}^d(E)^{\frac{1}{d}}} \le \frac{\nabla \cdot T(x)}{d}\frac{1}{\mathscr{L}^d(B)^{\frac{1}{d}}} \qquad \forall x \in E.$$

Integrating over E and applying the divergence theorem we get

$$\mathscr{L}^d(E)^{1-\frac{1}{d}} \le \frac{1}{d\mathscr{L}^d(B)^{1/d}}\int_E \nabla \cdot T(x)dx = \frac{1}{d\mathscr{L}^d(B)^{1/d}}\int_{\partial E}\langle T(x), \nu(x)\rangle \, d\mathcal{H}^{d-1}(x),$$

where $\nu : \partial E \to \mathbb{R}^d$ is the outer unit normal vector. Since $T(x) \in B$ for every $x \in E$, we have $|T(x)| \le 1$ for $x \in \partial E$ and thus $\langle T(x), \nu(x)\rangle \le 1$. We conclude with

$$\mathscr{L}^d(E)^{1-\frac{1}{d}} \le \frac{1}{d\mathscr{L}^d(B)^{1/d}}\int_{\partial E}\langle T(x), \nu(x)\rangle \, d\mathcal{H}^{d-1}(x) \le \frac{P(E)}{d\mathscr{L}^d(B)^{1/d}}.$$

4.3 Sobolev Inequality

The Sobolev inequality in $\mathbb{R}^d$ reads as:

$$\left(\int |f|^{p^*}\right)^{1/p^*} \le C(d,p)\left(\int |\nabla f|^p\right)^{1/p}, \qquad \forall f \in W^{1,p}(\mathbb{R}^d),$$

where $1 \le p < d$, $p^* := \frac{dp}{d-p}$ and $C(d,p)$ is a constant which depends only on the dimension d and the exponent p.

We will prove it via a method which closely resemble the one just used for the isoperimetric inequality. Again, we will neglect all the smoothness issues. Fix d, p and observe that without loss of generality we can assume $f \ge 0$ and $\int |f|^{p*} = 1$, so that our aim is to prove that

$$\left(\int |\nabla f|^p\right)^{1/p} \ge C, \tag{4.1}$$

for some constant C not depending on f. Fix once and for all a smooth, non negative function $g:\mathbb{R}^d\to\mathbb{R}$ satisfying $\int g=1$, define the probability measures

$$\mu := f^{p^*}\mathscr{L}^d, \qquad \nu := g\mathscr{L}^d,$$

and let T be the optimal transport map from μ to ν (w.r.t. the cost given by the distance squared). The change of variable formula gives

$$g(T(x)) = \frac{f^{p^*}(x)}{\det(\nabla T(x))}, \qquad \forall x\in\mathbb{R}^d.$$

Hence we have

$$\int g^{1-\frac{1}{d}} = \int g^{-\frac{1}{d}}g = \int (g\circ T)^{-\frac{1}{d}} f^{p^*} = \int \det(\nabla T)^{\frac{1}{d}} (f^{p^*})^{1-\frac{1}{d}}.$$

As for the case of the isoperimetric inequality, we know that T is the gradient of a convex function, thus $\nabla T(x)$ is a symmetric matrix with non negative eigenvalues and the arithmetic-geometric mean inequality gives $(\det(\nabla T(x)))^{1/d}\leq\frac{\nabla\cdot T(x)}{d}$. Thus we get

$$\int g^{1-\frac{1}{d}} \leq \frac{1}{d}\int \nabla\cdot T (f^{p^*})^{1-\frac{1}{d}} = -\frac{p^*}{d}\left(1-\frac{1}{d}\right)\int f^{\frac{p^*}{q}} T\cdot\nabla f,$$

where $\frac{1}{p}+\frac{1}{q}=1$. Finally, by Hölder inequality we have

$$\int g^{1-\frac{1}{d}} \leq \frac{p^*}{d}\left(1-\frac{1}{d}\right)\left(\int f^{p^*}|T|^q\right)^{\frac{1}{q}}\left(\int|\nabla f|^p\right)^{\frac{1}{p}} = \frac{p^*}{d}\left(1-\frac{1}{d}\right)\left(\int g(y)|y|^q dy\right)^{\frac{1}{q}}\left(\int|\nabla f|^p\right)^{\frac{1}{p}}.$$

Since g was a fixed given function, (4.1) is proved.

4.4 Bibliographical notes

The possibility of proving Brunn-Minkowski inequality via a change of variable is classical. It has been McCann in his PhD thesis [62] to notice that the use of optimal transport leads to a natural choice of reparametrization. It is interesting to notice that this approach can be generalized to curved and non-smooth spaces having *Ricci curvature bounded below*, see Proposition 7.14.

The idea of proving the isoperimetric inequality via a change of variable argument is due to Gromov [65]: in Gromov's proof it is not used the optimal transport map, but the so called Knothe's map. Such a map has the property that its gradient has non negative eigenvalues at every point, and the reader can easily check that this is all we used of Brenier's map in our proof, so that the argument of Gromov is the same we used here. The use of Brenier's map instead of Knothe's one makes the difference when studying the quantitative version of the isoperimetric problem: Figalli, Maggi and Pratelli in [38], using tools coming from optimal transport, proved the sharp quantitative isoperimetric inequality in $\mathbb{R}^d$ endowed with any norm (the sharp quantitative isoperimetric inequality for the Euclidean norm was proved earlier by Fusco, Maggi and Pratelli in [40] by completely different means).

The approach used here to prove the Sobolev inequality has been generalized by Cordero-Erasquin, Nazaret and Villani in [30] to provide a new proof of the sharp Gagliardo-Nirenberg-Sobolev inequality together with the identification of the functions realizing the equality

5 Variants of the Wasserstein distance

In this chapter we make a quick overview of some variants of the Wasserstein distance W_2 together with their applications. No proofs will be reported: our goal here is only to show that concepts coming from the transport theory can be adapted to cover a broader range of applications.

5.1 Branched optimal transportation

Consider the transport problem with $\mu := \delta_x$ and $\nu := \frac{1}{2}(\delta_{y_1} + \delta_{y_2})$ for the cost given by the distance squared on $\mathbb{R}^d$. Then Theorem 2.10 and Remark 2.13 tell that the unique geodesic (μ_t) connecting μ to ν is given by

$$\mu_t := \frac{1}{2}\Big(\delta_{(1-t)x+ty_1} + \delta_{(1-t)x+ty_2}\Big),$$

so that the geodesic produces a 'V-shaped' path.

For some applications, this is unnatural: for instance in real life networks, when one wants to transport the good located in x to the destinations y_1 and y_2 it is preferred to produce a branched structure, where first the good it is transported 'on a single truck' to some intermediate point, and only later split into two parts which are delivered to the 2 destinations. This produces a 'Y-shaped' path.

If we want to model the fact that 'it is convenient to ship things together', we are lead to the following construction, due to Gilbert. Say that the starting distribution of mass is given by $\mu = \sum_i a_i \delta_{x_i}$ and that the final one is $\nu = \sum_j b_j \delta_{y_j}$, with $\sum_i a_i = \sum_j b_j = 1$. An admissible dynamical transfer is then given by a finite, oriented, weighted graph G, where the weight is a function $w : \{set\ of\ edges\ of\ G\} \to \mathbb{R}$, satisfying the Kirchoff's rule:

$$\begin{aligned}
\sum_{\text{edges } e \text{ outgoing from } x_i} w(e) \quad &- \sum_{\text{edges } e \text{ incoming in } x_i} w(e) = \ a_i, \qquad \forall i\\
\sum_{\text{edges } e \text{ outgoing from } y_j} w(e) \quad &- \sum_{\text{edges } e \text{ incoming in } y_j} w(e) = -b_j, \qquad \forall j\\
\sum_{\text{edges } e \text{ outgoing from } z} w(e) \quad &- \sum_{\text{edges } e \text{ incoming in } z} w(e) = \ 0, \qquad \text{for any 'internal' node } z \text{ of } G
\end{aligned}$$

Then for $\alpha \in [0,1]$ one minimizes

$$\sum_{\text{edges } e \text{ of } G} w^\alpha(e) \cdot \text{length}(e),$$

among all admissible graphs G.

Observe that for $\alpha = 0$ this problem reduces to the classical Steiner problem, while for $\alpha = 1$ it reduces to the classical optimal transport problem for $cost = distance$.

It is not hard to show the existence of a minimizer for this problem. What is interesting, is that a 'continuous' formulation is possible as well, which allows to discuss the minimization problem for general initial and final measure in $\mathscr{P}(\mathbb{R}^d)$.

Definition 5.1 (Admissible continuous dynamical transfer) *Let $\mu, \nu \in \mathscr{P}(\mathbb{R}^d)$. An admissible continuous dynamical transfer from μ to ν is given by a countably $\mathcal{H}^1$-rectifiable set Γ, an orientation on it $\tau : \Gamma \to S^{d-1}$, and a weight function $w : \Gamma \to [0, +\infty)$, such that the $\mathbb{R}^d$ valued measure $J_{\Gamma,\tau,w}$ defined by*

$$J_{\Gamma,\tau,w} := w\tau\mathcal{H}^1|_\Gamma,$$

satisfies

$$\nabla \cdot J_{\Gamma,\tau,w} = \nu - \mu,$$

(which is the natural generalization of the Kirchoff rule).

Given $\alpha \in [0,1]$, the cost function associated to (Γ, τ, w) is defined as

$$\mathcal{E}_\alpha(J_{\Gamma,\tau,w}) := \int_\Gamma w^\alpha \, d\mathscr{H}^1.$$

Theorem 5.2 (Existence) *Let μ, $\nu \in \mathscr{P}(\mathbb{R}^d)$ with compact support. Then for all $\alpha \in [0,1)$ there exists a minimizer of the cost in the set of admissible continuous dynamical transfers connecting μ to ν. If $\mu = \delta_z$ and $\nu = \mathcal{L}^d|_{[0,1]^d}$, the minimal cost is finite if and only if $\alpha > 1 - 1/d$.*

The fact that $1 - 1/d$ is a limit value to get a finite cost, can be heuristically understood by the following calculation. Suppose we want to move a Delta mass δ_x into the Lebesgue measure on a unit cube whose center is x. Then the first thing one wants to try is: divide the cube into 2^d cubes of side length $1/2$, then split the delta into 2^d masses and let them move onto the centers of these 2^d cubes. Repeat the process by dividing each of the 2^d cubes into 2^d cubes of side length $1/4$ and so on. The total cost of this dynamical transfer is proportional to:

$$\sum_{i=1}^{\infty} \underbrace{2^{id}}_{\substack{\text{number of segments}\\ \text{at the step } i}} \underbrace{\frac{1}{2^i}}_{\substack{\text{length of each}\\ \text{segment at the step } i}} \underbrace{\frac{1}{2^{\alpha id}}}_{\substack{\text{weighted mass on each}\\ \text{segment at the step } i}} = \sum_{i=1}^{\infty} 2^{i(d-1-\alpha d)},$$

which is finite if and only if $d - 1 - \alpha d < 0$, that is, if and only if $\alpha > 1 - \frac{1}{d}$.

A regularity result holds for $\alpha \in (1 - 1/d, 1)$ which states that far away from the supports of the starting and final measures, any minimal transfer is actually a finite tree:

Theorem 5.3 (Regularity) *Let $\mu, \nu \in \mathscr{P}(\mathbb{R}^d)$ with compact support, $\alpha \in (1 - 1/n, 1)$ and let (Γ, τ, w) be a continuous tree with minimal α-cost between μ and ν. Then Γ is locally a finite tree in $\mathbb{R}^d \setminus (\operatorname{supp} \mu \cup \operatorname{supp} \nu)$.*

5.2 Different action functional

Let us recall that the Benamou-Brenier formula (Proposition 2.30) identifies the squared Wasserstein distance between $\mu^0 = \rho^0 \mathscr{L}^d, \mu^1 := \rho^1 \mathscr{L}^d \in \mathscr{P}_2(\mathbb{R}^d)$ by

$$W_2^2(\mu^0, \mu^1) = \inf \int_0^1 \int |v_t|^2(x) \rho_t(x) d\mathscr{L}^d(x) dt,$$

where the infimum is taken among all the distributional solutions of the continuity equation

$$\frac{d}{dt}\rho_t + \nabla \cdot (v_t \rho_t) = 0,$$

with $\rho_0 = \rho^0$ and $\rho_1 = \mu^1$.

A natural generalization of the distance W_2 comes by considering a new action, modified by putting a weight on the density, that is: given a smooth function $h : [0,\infty) \to [0,\infty)$ we define

$$W_h^2(\rho^0 \mathcal{L}^d, \rho^1 \mathcal{L}^d) = \inf \int_0^1 \int |v_t|^2(x) h(\rho_t(x)) d\mathscr{L}^d(x) dt, \tag{5.1}$$

where the infimum is taken among all the distributional solutions of the *non linear* continuity equation

$$\frac{d}{dt}\rho_t + \nabla\cdot(v_t h(\rho_t)) = 0, \tag{5.2}$$

with $\rho_0 = \rho^0$ and $\rho_1 = \rho^1$.

The key assumption that leads to the existence of an action minimizing curve is the concavity of h, since this leads to the joint convexity of

$$(\rho, J) \mapsto h(\rho)\left|\frac{J}{h(\rho)}\right|^2,$$

so that using this convexity with $J = vh(\rho)$, one can prove existence of minima of (5.1). Particularly important is the case given by $h(z) := z^\alpha$ for $\alpha < 1$ from which we can build the distance $\tilde{W}_\alpha$ defined by

$$\tilde{W}_\alpha(\rho^0\mathscr{L}^d, \rho^1\mathscr{L}^d) := \left(\inf\int_0^1\int |v_t|^2(x)\rho_t^{2-\alpha}(x)d\mathscr{L}^d(x)dt\right)^{\frac{1}{\alpha}}, \tag{5.3}$$

the infimum being taken among all solutions of (5.2) with $\rho_0 = \rho^0$ and $\rho_1 = \rho^1$. The following theorem holds:

Theorem 5.4 *Let $\alpha > 1 - \frac{1}{d}$. Then the infimum in* (5.3) *is always reached and, if it is finite, the minimizer is unique. Now fix a measure $\mu \in \mathscr{P}(\mathbb{R}^d)$. The set of measures ν with $\tilde{W}_\alpha(\mu,\nu) < \infty$ endowed with $\tilde{W}_\alpha$ is a complete metric space and bounded subsets are narrowly compact.*

We remark that the behavior of action minimizing curves in this setting is, in some very rough sense, "dual" of the behavior of the branched optimal transportation discussed in the previous section. Indeed, in this problem the mass tends to spread out along an action minimizing curve, rather than to glue together.

5.3 An extension to measures with unequal mass

Let us come back to the Heat equation seen as Gradient Flow of the entropy functional $E(\rho) = \int \rho\log(\rho)$ with respect to the Wasserstein distance W_2, as discussed at the beginning of Section 3.3 and in Subsection 3.3.2. We discussed the topic for arbitrary probability measures in $\mathbb{R}^d$, but actually everything could have been done for probability measures concentrated on some open bounded set $\Omega \subset \mathbb{R}^d$ with smooth boundary, that is: consider the metric space $(\mathscr{P}(\Omega), W_2)$ and the entropy functional $E(\rho) = \int \rho\log(\rho)$ for absolutely continuous measures and $E(\mu) = +\infty$ for measures with a singular part. Now use the Minimizing Movements scheme to build up a family of discrete solutions ρ_t^τ starting from some given measure $\overline{\rho} \in \mathscr{P}(\Omega)$. It is then possible to see that these discrete families converge as $\tau \downarrow 0$ to the solution of the Heat equation with *Neumann boundary condition*:

$$\begin{cases} \frac{d}{dt}\rho_t = \Delta\rho_t, & \text{in } \Omega\times(0,+\infty), \\ \rho_t \to \overline{\rho}, & \text{weakly as } t\to 0 \\ \nabla\rho_t\cdot\eta = 0, & \text{in } \partial\Omega\times(0,\infty), \end{cases}$$

where η is the outward pointing unit vector on $\partial\Omega$.

The fact that the boundary condition is the Neumann's one, can be heuristically guessed by the fact that working in $\mathscr{P}(\Omega)$ enforces the mass to be constant, with no flow of the mass through the boundary.

It is then natural to ask whether it is possible to modify the transportation distance in order to take into account measures with unequal masses, and such that the Gradient Flow of the entropy functional produces solutions of the Heat equation in Ω with Dirichlet boundary conditions. This is actually doable, as we briefly discuss now.

Let $\Omega \subset \mathbb{R}^d$ be open and bounded. Consider the set $\mathcal{M}_2(\Omega)$ defined by

$$\mathcal{M}_2(\Omega) := \Big\{ \text{measures } \mu \text{ on } \Omega \text{ such that } \int d^2(x, \partial\Omega) d\mu(x) < \infty \Big\},$$

and for any $\mu, \nu \in \mathcal{M}_2(\Omega)$ define the set of admissible transfer plans $\mathrm{Adm}_b(\mu,\nu)$ by: $\boldsymbol{\gamma} \in \mathrm{Adm}_b(\mu,\nu)$ if and only if $\boldsymbol{\gamma}$ is a measure on $(\overline{\Omega})^2$ such that

$$\pi^1_\# \boldsymbol{\gamma}_{|\Omega} = \mu, \qquad \pi^2_\# \boldsymbol{\gamma}_{|\Omega} = \nu.$$

Notice the difference w.r.t. the classical definition of transfer plan: here we are requiring the first (respectively, second) marginal to coincide with μ (respectively ν) only inside the open set Ω. This means that in transferring the mass from μ to ν we are free to take/put as much mass as we want from/to the boundary. Then one defines the *cost* $C(\boldsymbol{\gamma})$ of a plan $\boldsymbol{\gamma}$ by

$$C(\boldsymbol{\gamma}) := \int |x-y|^2 d\boldsymbol{\gamma}(x,y),$$

and then the distance Wb_2 by

$$Wb_2(\mu,\nu) := \inf \sqrt{C(\boldsymbol{\gamma})},$$

where the infimum is taken among all $\boldsymbol{\gamma} \in \mathrm{Adm}_b(\mu,\nu)$.

The distance Wb_2 shares many properties with the Wasserstein distance W_2.

Theorem 5.5 (Main properties of Wb_2) *The following hold:*

- *Wb_2 is a distance on $\mathcal{M}_2(\Omega)$ and the metric space $(\mathcal{M}_2(\Omega), Wb_2)$ is Polish and geodesic.*
- *A sequence $(\mu_n) \subset \mathcal{M}_2(\Omega)$ converges to μ w.r.t. Wb_2 if and only if μ_n converges weakly to μ in duality with continuous functions with compact support in Ω and $\int d^2(x,\partial\Omega)d\mu_n \to \int d^2(x,\partial\Omega) d\mu$ as $n \to \infty$.*
- *Finally, a plan $\boldsymbol{\gamma} \in Adm_b(\mu,\nu)$ is optimal (i.e. it attains the minimum cost among admissible plans) if and only there exists a c-concave function φ* which is identically 0 on $\partial\Omega$ *such that* $\mathrm{supp}(\boldsymbol{\gamma}) \subset \partial^c\varphi$ *(here $c(x,y) = |x-y|^2$).*

Observe that $(\mathcal{M}_2(\Omega), Wb_2)$ is always a geodesic space (while from Theorem 2.10 and Remark 2.14 we know that $(\mathscr{P}(\Omega), W_2)$ is geodesic if and only if Ω is, that is, if and only if Ω is convex).

It makes perfectly sense to extend the entropy functional to the whole $\mathcal{M}_2(\Omega)$: the formula is still $E(\mu) = \int \rho \log(\rho)$ for $\mu = \rho \mathcal{L}^d{}_{|\Omega}$, and $E(\mu) = +\infty$ for measures not absolutely continuous. The Gradient Flow of the entropy w.r.t. Wb_2 produces solutions of the Heat equation with Dirichlet boundary conditions in the following sense:

Theorem 5.6 *Let $\mu \in \mathcal{M}_2(\Omega)$ be such that $E(\mu) < \infty$. Then:*

- *for every $\tau > 0$ there exists a unique discrete solution ρ^τ_t starting from μ and constructed via the Minimizing Movements scheme as in Definition 3.7.*
- *As $\tau \downarrow 0$, the measures ρ^τ_t converge to a unique measure ρ_t in $(\mathcal{M}_2(\Omega), Wb_2)$ for any $t > 0$.*
- *The map $(x,t) \mapsto \rho_t(x)$ is a solution of the Heat equation*

$$\begin{cases} \frac{d}{dt}\rho_t &= \Delta \rho_t, & \text{in } \Omega \times (0,+\infty), \\ \quad \rho_t &\to \mu, & \text{weakly as } t \to 0, \end{cases}$$

subject to the Dirichlet boundary condition $\rho_t(x) = e^{-1}$ in $\partial\Omega$ for every $t > 0$ (that is, $\rho_t - e^{-1}$ belongs to $H^1_0(\Omega)$ for every $t > 0$).

The fact that the boundary value is given by e^{-1} can be heuristically guessed by the fact that the entropy has a global minimum in $\mathcal{M}_2(\Omega)$: such minimum is given by the measure with constant density e^{-1}, i.e. the measure whose density is everywhere equal to the minimum of $z \mapsto z\log(z)$.

On the bad side, the entropy E is *not* geodesically convex in $(\mathcal{M}_2(\Omega), Wb_2)$, and this implies that it is not clear whether the strong properties of Gradient Flows w.r.t. W_2 as described in Section 3.3 - Theorem 3.35 and Proposition 3.38 are satisfied also in this setting. In particular, it is not clear whether there is contractivity of the distance or not:

Open Problem 5.7 *Let ρ_t^1, ρ_t^2 two solutions of the Heat equation with Dirichlet boundary condition $\rho_t^i = e^{-1}$ in $\partial\Omega$ for every $t > 0$, $i = 1, 2$. Prove or disprove that*

$$Wb_2(\rho_s^1, \rho_s^2) \le Wb_2(\rho_t^1, \rho_t^2), \qquad \forall t > s.$$

The question is open also for convex and smooth open sets Ω.

5.4 Bibliographical notes

The connection of branched transport and transport problem as discussed in Section 5.1 was first pointed out by Q. Xia in [81]. An equivalent model was proposed by F. Maddalena, J.-M. Morel and S. Solimini in [61]. In [81], [60] and [15] the existence of an optimal branched transport (Theorem 5.2) was also provided. Later, this result has been extended in several directions, see for instance the works A. Brancolini, G. Buttazzo and F. Santambrogio ([16]) and Bianchini-Brancolini [15]. The interior regularity result (Theorem 5.3) has been proved By Q. Xia in [82] and M. Bernot, V. Caselles and J.-M. Morel in [14]. Also, we remark that L. Brasco, G. Buttazzo and F. Santambrogio proved a kind of Benamou-Brenier formula for branched transport in [17].

The content of Section 5.2 comes from J. Dolbeault, B. Nazaret and G. Savaré [33] and [26] of J. Carrillo, S. Lisini, G. Savaré and D. Slepcev.

Section 5.3 is taken from a work of the second author and A. Figalli [37].

6 More on the structure of $(\mathscr{P}_2(M), W_2)$

The aim of this Chapter is to give a comprehensive description of the structure of the 'Riemannian manifold' $(\mathscr{P}_2(\mathbb{R}^d), W_2)$, thus the content of this part of the work is the natural continuation of what we discussed in Subsection 2.3.2. For the sake of simplicity, we are going to stick to the Wasserstein space on $\mathbb{R}^d$, but the reader should keep in mind that the discussions here can be generalized with only little effort to the Wasserstein space built over a Riemannian manifold.

6.1 "Duality" between the Wasserstein and the Arnold Manifolds

The content of this section is purely formal and directly comes from the seminal paper of Otto [67]. We won't even try to provide a rigorous background for the discussion we will do here, as we believe that dealing with the technical problems would lead the reader far from the geometric intuition. Also, we will not use the "results" presented here later on: we just think that these concepts are worth of mention. Thus for the purpose of this section just think that 'each measure is absolutely continuous with smooth density', that 'each L^2 function is C^∞', and so on.

Let us recall the definition of Riemannian submersion. Let M, N be Riemannian manifolds and let $f : M \to N$ a smooth map. f is a submersion provided the map:

$$df : \mathrm{Ker}^\perp\big(df(x)\big) \to T_{f(x)}N,$$

is a surjective isometry for any $x \in M$. A trivial example of submersion is given in the case $M = N \times L$ (for some Riemannian manifold L, with M endowed with the product metric) and $f : M \to N$ is the natural projection. More generally, if f is a Riemannian submersion, for each $y \in N$, the set $f^{-1}(y) \subset M$ is a smooth Riemannian submanifold.

The "duality" between the Wasserstein and the Arnold Manifolds consists in the fact that there exists a Big Manifold BM which is flat and a natural Riemannian submersion from BM to $\mathscr{P}_2(\mathbb{R}^d)$ whose fibers are precisely the Arnold Manifolds.

Let us define the objects we are dealing with. Fix once and for all a reference measure $\overline{\rho} \in \mathscr{P}_2(\mathbb{R}^d)$ (recall that we are "assuming" that all the measures are absolutely continuous with smooth densities - so that we will use the same notation for both the measure and its density).

- The Big Manifold BM is the space $L^2(\overline{\rho})$ of maps from $\mathbb{R}^d$ to $\mathbb{R}^d$ which are L^2 w.r.t. the reference measure $\overline{\rho}$. The tangent space at some map $T \in \mathrm{BM}$ is naturally given by the set of vector fields belonging to $L^2(\overline{\rho})$, where the perturbation of T in the direction of the vector field u is given by $t \mapsto T + tu$.
- The target manifold of the submersion is the Wasserstein "manifold" $\mathscr{P}_2(\mathbb{R}^d)$. We recall that the tangent space $\mathrm{Tan}_\rho(\mathscr{P}_2(\mathbb{R}^d))$ at the measure ρ is the set

  $$\mathrm{Tan}_\rho(\mathscr{P}_2(\mathbb{R}^d)) := \Big\{ \nabla\varphi \ : \ \varphi \in C_c^\infty(\mathbb{R}^d) \Big\},$$

  endowed with the scalar product of $L^2(\rho)$ (we neglect to take the closure in $L^2(\rho)$ because we want to keep the discussion at a formal level). The perturbation of a measure ρ in the direction of a tangent vector $\nabla\varphi$ is given by $t \mapsto (Id + t\nabla\varphi)_\#\rho$.
- The Arnold Manifold $\mathrm{Arn}(\rho)$ associated to a certain measure $\rho \in \mathscr{P}_2(\mathbb{R}^d)$ is the set of maps $S : \mathbb{R}^d \to \mathbb{R}^d$ which preserve ρ:

  $$\mathrm{Arn}(\rho) := \Big\{ S : \mathbb{R}^d \to \mathbb{R}^d \ : \ S_\#\rho = \rho \big\}.$$

  We endow $\mathrm{Arn}(\rho)$ with the L^2 distance calculated w.r.t. ρ. To understand who is the tangent space at $\mathrm{Arn}(\rho)$ at a certain map S, pick a vector field v on $\mathbb{R}^d$ and consider the perturbation $t \mapsto S + tv$ of S in the direction of v. Then v is a tangent vector if and only if $\frac{d}{dt}|_{t=0}(S + tv)_\#\rho = 0$. Observing that

  $$\frac{d}{dt}|_{t=0}(S+tv)_\#\rho = \frac{d}{dt}|_{t=0}(Id+tv\circ S^{-1})_\#(S_\#\rho) = \frac{d}{dt}|_{t=0}(Id+tv\circ S^{-1})_\#\rho = \nabla\cdot(v\circ S^{-1}\rho),$$

  we deduce

  $$\mathrm{Tan}_S\mathrm{Arn}(\rho) = \Big\{ \text{vector fields } v \text{ on } \mathbb{R}^d \text{ such that } \nabla \cdot (v \circ S^{-1}\rho) = 0 \Big\},$$

  which is naturally endowed with the scalar product in $L^2(\rho)$.

  We are calling the manifold $\mathrm{Arn}(\rho)$ an Arnold Manifold, because if ρ is the Lebesgue measure restricted to some open, smooth and bounded set Ω, this definition reduces to the well known definition of Arnold manifold in fluid mechanics: the geodesic equation in such space is - formally - the Euler equation for the motion of an incompressible and inviscid fluid in Ω.
- Finally, the "Riemannian submersion" Pf from BM to $\mathscr{P}_2(\mathbb{R}^d)$ is the push forward map:

  $$\begin{array}{rcl} \mathrm{Pf} : \mathrm{BM} & \to & \mathscr{P}_2(\mathbb{R}^d), \\ T & \mapsto & T_\#\overline{\rho}, \end{array}$$

We claim that Pf is a Riemannian submersion and that the fiber $\mathrm{Pf}^{-1}(\rho)$ is isometric to the manifold $\mathrm{Arn}(\rho)$.

We start considering the fibers. Fix $\rho \in \mathscr{P}_2(\mathbb{R}^d)$. Observe that

$$\mathrm{Pf}^{-1}(\rho) = \Big\{ T \in \mathrm{BM} \ : \ T_{\#}\overline{\rho} = \rho \Big\},$$

and that the tangent space $\mathrm{Tan}_T \mathrm{Pf}^{-1}(\rho)$ is the set of vector fields u such that $\frac{d}{dt}|_{t=0}(T+tu)_{\#}\overline{\rho} = 0$, so that from

$$\frac{d}{dt}|_{t=0}(T+tu)_{\#}\overline{\rho} = \frac{d}{dt}|_{t=0}(Id+tu\circ T^{-1})_{\#}(T_{\#}\overline{\rho}) = \frac{d}{dt}|_{t=0}(Id+tu\circ T^{-1})_{\#}\rho = \nabla\cdot(u\circ T^{-1}\rho),$$

we have

$$\mathrm{Tan}_T \mathrm{Pf}^{-1}(\rho) = \Big\{ \text{vector fields } u \text{ on } \mathbb{R}^d \text{ such that } \nabla\cdot(u\circ T^{-1}\rho) = 0 \Big\},$$

and the scalar product between two vector fields in $\mathrm{Tan}_T \mathrm{Pf}^{-1}(\rho)$ is the one inherited by the one in BM, i.e. is the scalar product in $L^2(\overline{\rho})$.

Now choose a distinguished map $T^\rho \in \mathrm{Pf}^{-1}(\rho)$ and notice that the right composition with T^ρ provides a natural bijective map from $\mathrm{Arn}(\rho)$ into $\mathrm{Pf}^{-1}(\rho)$, because

$$S_{\#}\rho = \rho \qquad \Leftrightarrow \qquad (S\circ T^\rho)_{\#}\overline{\rho} = \rho.$$

We claim that this right composition also provides an isometry between the "Riemannian manifolds" $\mathrm{Arn}(\rho)$ and $\mathrm{Pf}^{-1}(\rho)$: indeed, if $v \in \mathrm{Tan}_S \mathrm{Arn}(\rho)$, then the perturbed maps $S + tv$ are sent to $S\circ T^\rho + tv\circ T^\rho$, which means that the perturbation v of S is sent to the perturbation $u := v\circ T^\rho$ of $S\circ T^\rho$ by the differential of the right composition. The conclusion follows from the change of variable formula, which gives

$$\int |v|^2 d\rho = \int |u|^2 d\overline{\rho}.$$

Clearly, the kernel of the differential $d\mathrm{Pf}$ of Pf at T is given by $\mathrm{Tan}_T \mathrm{Pf}^{-1}\big(\mathrm{Pf}(T)\big)$, thus it remains to prove that its orthogonal is sent isometrically onto $\mathrm{Tan}_{\mathrm{Pf}(T)}(\mathscr{P}_2(\mathbb{R}^d))$ by $d\mathrm{Pf}$. Fix $T \in \mathrm{BM}$, let $\rho := \mathrm{Pf}(T) = T_{\#}\overline{\rho}$ and observe that

$$\begin{aligned}
\mathrm{Tan}_T^{\perp}\big(\mathrm{Pf}^{-1}(\rho)\big) &= \Big\{ \text{vector fields } w \ : \ \int \langle w, u\rangle \, d\overline{\rho} = 0, \ \forall u \text{ s.t. } \nabla\cdot(u\circ T^{-1}\rho) = 0 \Big\} \\
&= \Big\{ \text{vector fields } w \ : \ \int \big\langle w\circ T^{-1}, u\circ T^{-1}\big\rangle \, d\rho = 0, \ \forall u \text{ s.t. } \nabla\cdot(u\circ T^{-1}\rho) = 0 \Big\} \\
&= \Big\{ \text{vector fields } w \ : \ w\circ T^{-1} = \nabla\varphi \text{ for some } \varphi \in C_c^\infty(\mathbb{R}^d) \Big\}.
\end{aligned}$$

Now pick $w \in \mathrm{Tan}_T^{\perp}\big(\mathrm{Pf}^{-1}(\rho)\big)$, let $\varphi \in C_c^\infty(\mathbb{R}^d)$ be such that $w\circ T^{-1} = \nabla\varphi$ and observe that

$$\frac{d}{dt}|_{t=0}\mathrm{Pf}(T+tw) = \frac{d}{dt}|_{t=0}(T+tw)_{\#}\overline{\rho} = \frac{d}{dt}|_{t=0}(Id+tw\circ T^{-1})_{\#}(T_{\#}\overline{\rho}) = \frac{d}{dt}|_{t=0}(Id+t\nabla\varphi)_{\#}\rho,$$

which means, by definition of $\mathrm{Tan}_\rho(\mathscr{P}_2(\mathbb{R}^d))$ and the action of tangent vectors, that the differential $d\mathrm{Pf}(T)(w)$ of Pf calculated at T along the direction w is given by $\nabla\varphi$. The fact that this map is an isometry follows once again by the change of variable formula

$$\int |w|^2 d\overline{\rho} = \int |w\circ T^{-1}|^2 d\rho = \int |\nabla\varphi|^2 d\rho.$$

6.2 On the notion of tangent space

Aim of this section is to quickly discuss the definition of tangent space of $\mathscr{P}_2(\mathbb{R}^d)$ at a certain measure μ from a purely geometric perspective. We will see how this perspective is related to the discussion made in Subsection 2.3.2, where we defined tangent space as

$$\mathrm{Tan}_\mu(\mathscr{P}_2(\mathbb{R}^d)) := \overline{\left\{\nabla\varphi\ :\ \varphi \in C_c^\infty(\mathbb{R}^d)\right\}}^{L^2(\mathbb{R}^d,\mathbb{R}^d;\mu)}.$$

Recall that this definition came from the characterization of absolutely continuous curves on $\mathscr{P}_2(\mathbb{R}^d)$ (Theorem 2.29 and the subsequent discussion).

Yet, there is a completely different and purely geometrical approach which leads to a definition of tangent space at μ. The idea is to think the tangent space at μ as the "space of directions", or, which is the same, as the set of constant speed geodesics emanating from μ. More precisely, let the set $\mathcal{Geod}_\mu$ be defined by:

$$\mathcal{Geod}_\mu := \left\{ \begin{array}{l} \text{constant speed geodesics starting from } \mu \\ \text{and defined on some interval of the kind } [0,T] \end{array} \right\} / \approx,$$

where we say that $(\mu_t) \approx (\mu'_t)$ provided they coincide on some right neighborhood of 0. The natural distance D on $\mathcal{Geod}_\mu$ is:

$$D\big((\mu_t),(\mu'_t)\big) := \varlimsup_{t\downarrow 0} \frac{W_2(\mu_t,\mu'_t)}{t}. \tag{6.1}$$

The *Geometric Tangent space* $\mathbf{Tan}_\mu(\mathscr{P}_2(\mathbb{R}^d))$ is then defined as the completion of $\mathcal{Geod}_\mu$ w.r.t. the distance D.

The natural question here is: what is the relation between the "space of gradients" $\mathrm{Tan}_\mu(\mathscr{P}_2(\mathbb{R}^d))$ and the "space of directions" $\mathbf{Tan}_\mu(\mathscr{P}_2(\mathbb{R}^d))$?

Recall that from Remark 1.22 we know that given $\varphi \in C_c^\infty(\mathbb{R}^d)$, the map $t \mapsto (Id + t\nabla\varphi)_\#\mu$ is a constant speed geodesic on a right neighborhood of 0. This means that there is a natural map ι_μ from the set $\{\nabla\varphi : \varphi \in C_c^\infty\}$ into $\mathcal{Geod}_\mu$, and therefore into $\mathbf{Tan}_\mu(\mathscr{P}_2(\mathbb{R}^d))$, which sends $\nabla\varphi$ into the (equivalence class of the) geodesic $t \mapsto (Id + t\nabla\varphi)_\#\mu$. The main properties of the Geometric Tangent space and of this map are collected in the following theorem, which we state without proof.

Theorem 6.1 (The tangent space) *Let $\mu \in \mathscr{P}_2(\mathbb{R}^d)$. Then:*

- *the $\varlimsup$ in (6.1) is always a limit,*
- *the metric space $(\mathbf{Tan}_\mu(\mathscr{P}_2(\mathbb{R}^d)), D)$ is complete and separable,*
- *the map $\iota_\mu : \{\nabla\varphi\} \to \mathbf{Tan}_\mu(\mathscr{P}_2(\mathbb{R}^d))$ is an injective isometry, where on the source space we put the L^2 distance w.r.t. μ. Thus, ι_μ always extends to a natural isometric embedding of $\mathrm{Tan}_\mu(\mathscr{P}_2(\mathbb{R}^d))$ into $\mathbf{Tan}_\mu(\mathscr{P}_2(\mathbb{R}^d))$.*

Furthermore, the following statements are equivalent:

i) the space $(\mathbf{Tan}_\mu(\mathscr{P}_2(\mathbb{R}^d)), D)$ is an Hilbert space,

ii) the map $\iota_\mu : \mathrm{Tan}_\mu(\mathscr{P}_2(\mathbb{R}^d)) \to \mathbf{Tan}_\mu(\mathscr{P}_2(\mathbb{R}^d))$ is surjective,

iii) the measure μ is regular (definition 1.25).

We comment on the second part of the theorem. The first thing to notice is that the "space of directions" $\mathbf{Tan}_\mu(\mathscr{P}_2(\mathbb{R}^d))$ can be strictly larger than 'the space of gradients' $\mathrm{Tan}_\mu(\mathscr{P}_2(\mathbb{R}^d))$. This is actually not surprising if one thinks to the case in which μ is a Dirac mass. Indeed in this situation the space $(\mathbf{Tan}_\mu(\mathscr{P}_2(\mathbb{R}^d)), D)$ coincides with the space $(\mathscr{P}_2(\mathbb{R}^d), W_2)$ (this can be checked directly from the definition), however, the space $\mathrm{Tan}_\mu(\mathscr{P}_2(\mathbb{R}^d))$ is actually isometric to $\mathbb{R}^d$ itself, and is therefore much smaller.

The reason is that geodesics are not always induced by maps, that is, they are not always of the form $t \mapsto (Id + tu)_\# \mu$ for some vector field $u \in L^2_\mu$. To some extent, here we are facing the same problem we had to face when starting the study of the optimal transport problem: maps are typically not sufficient to produce (optimal) transports. From this perspective, it is not surprising that if the measure we are considering is regular (that is, if for any $\nu \in \mathscr{P}_2(\mathbb{R}^d)$ there exists a unique optimal plan, and this plan is induced by a map), then the "space of directions" coincides with the "space of directions induced by maps".

6.3 Second order calculus

Now we pass to the description of the second order analysis over $\mathscr{P}_2(\mathbb{R}^d)$. The concepts that now enter into play are: Covariant Derivative, Parallel Transport and Curvature. To some extent, the situation is similar to the one we discussed in Subsection 2.3.2 concerning the first order structure: the metric space $(\mathscr{P}_2(\mathbb{R}^d), W_2)$ is not a Riemannian manifold, but if we are careful in giving definitions and in the regularity requirements of the objects involved we will be able to perform calculations very similar to those valid in a genuine Riemannian context.

Again, we are restricting the analysis to the Euclidean case only for simplicity: all of what comes next can be generalized to the analysis over $\mathscr{P}_2(M)$, for a generic Riemannian manifold M.

On a typical course of basic Riemannian geometry, one of the first concepts introduced is that of Levi-Civita connection, which identifies the only natural ("natural" here means: "compatible with the Riemannian structure") way of differentiating vector fields on the manifold. It would therefore be natural to set up our discussion on the second order analysis on $\mathscr{P}_2(\mathbb{R}^d)$ by giving the definition of Levi-Civita connection in this setting. However, this cannot be done. The reason is that we don't have a notion of smoothness for vector fields, therefore not only we don't know how to covariantly differentiate vector fields, but we don't know either which are the vector fields regular enough to be differentiated. In a purely Riemannian setting this problem does not appear, as a Riemannian manifold borns as smooth manifold on which we define a scalar product on each tangent space; but the space $\mathscr{P}_2(\mathbb{R}^d)$ does not have a smooth structure (there is no diffeomorphism of a small ball around the origin in $\mathrm{Tan}_\mu(\mathscr{P}_2(\mathbb{R}^d))$ onto a neighborhood of μ in $\mathscr{P}_2(\mathbb{R}^d)$). Thus, we have to proceed in a different way, which we describe now:

Regular curves first of all, we drop the idea of defining a smooth vector field on the whole "manifold". We will rather concentrate on finding an appropriate definition of smoothness for vector fields defined along curves. We will see that to do this, we will need to work with a particular kind of curves, which we call *regular*, see Definition 6.2.

Smoothness of vector fields. We will then be able to define the smoothness of vector fields defined along regular curves (Definition 6.5). Among others, a notion of smoothness of particular relevance is that of *absolutely continuous* vector fields: for this kind of vector fields we have a natural notion of *total derivative* (not to be confused with the covariant one, see Definition 6.6).

Levi-Civita connection. At this point we have all the ingredients we need to define the covariant derivative and to prove that it is the Levi-Civita connection on $\mathscr{P}_2(\mathbb{R}^d)$ (Definiton 6.8 and discussion thereafter).

Parallel transport. This is the main existence result on this subject: we prove that along regular curves the parallel transport always exists (Theorem 6.15). We will also discuss a counterexample to the existence of parallel transport along a non-regular geodesic (Example 6.16). This will show that the definition of regular curve is not just operationally needed to provide a definition of smoothness

of vector fields, but is actually intrinsically related to the geometry of $\mathscr{P}_2(\mathbb{R}^d)$.
Calculus of derivatives. Using the technical tools developed for the study of the parallel transport, we will be able to explicitly compute the total and covariant derivatives of basic examples of vector fields.
Curvature. We conclude the discussion by showing how the concepts developed can lead to a rigorous definition of the curvature tensor on $\mathscr{P}_2(\mathbb{R}^d)$.

We will write $\|v\|_\mu$ and $\langle v, w\rangle_\mu$ for the norm of the vector field v and the scalar product of the vector fields v, w in the space $L^2(\mu)$ (which we will denote by L^2_μ), respectively.

We now start with the definition of regular curve. All the curves we will consider are defined on $[0,1]$, unless otherwise stated.

Definition 6.2 (Regular curve) *Let (μ_t) be an absolutely continuous curve and let (v_t) be its velocity vector field, that is (v_t) is the unique vector field - up to equality for a.e. t - such that $v_t \in \mathrm{Tan}_{\mu_t}(\mathscr{P}_2(\mathbb{R}^d))$ for a.e. t and the continuity equation*

$$\frac{d}{dt}\mu_t + \nabla \cdot (v_t \mu_t) = 0,$$

holds in the sense of distributions (recall Theorem 2.29 and Definition 2.31). We say that (μ_t) is regular provided

$$\int_0^1 \|v_t\|^2_{\mu_t} dt < \infty, \tag{6.2}$$

and

$$\int_0^1 \mathrm{Lip}(v_t) dt < \infty. \tag{6.3}$$

Observe that the validity of (6.3) is independent on the parametrization of the curve, thus if it is fulfilled it is always possible to reparametrize the curve (e.g. with constant speed) in order to let it satisfy also (6.2).

Now assume that (μ_t) is regular. Then by the classical Cauchy-Lipschitz theory we know that there exists a unique family of maps $\mathbf{T}(t,s,\cdot) : \mathrm{supp}(\mu_t) \to \mathrm{supp}(\mu_s)$ satisfying

$$\left\{ \begin{array}{rcll} \dfrac{d}{ds}\mathbf{T}(t,s,x) &=& v_s(\mathbf{T}(t,s,x)), & \forall t \in [0,1],\ x \in \mathrm{supp}(\mu_t),\ a.e.\ s \in [0,1], \\ \mathbf{T}(t,t,x) &=& x, & \forall t \in [0,1],\ x \in \mathrm{supp}(\mu_t). \end{array} \right. \tag{6.4}$$

Also it is possible to check that these maps satisfy the additional properties

$$\begin{array}{rcll} \mathbf{T}(r,s,\cdot) \circ \mathbf{T}(t,r,\cdot) &=& \mathbf{T}(t,s,\cdot) & \forall t,r,s \in [0,1], \\ \mathbf{T}(t,s,\cdot)_\# \mu_t &=& \mu_s, & \forall t,s \in [0,1]. \end{array}$$

We will call this family of maps the *flow maps* of the curve (μ_t). Observe that for any couple of times $t,s \in [0,1]$, the right composition with $\mathbf{T}(t,s,\cdot)$ provides a bijective isometry from $L^2_{\mu_s}$ to $L^2_{\mu_t}$. Also, notice that from condition (6.2) and the inequalities

$$\begin{aligned} \|\mathbf{T}(t,s,\cdot) - \mathbf{T}(t,s',\cdot)\|^2_{\mu_t} &\le \int \left(\int_s^{s'} v_r(\mathbf{T}(t,r,x)) dr \right)^2 d\mu_t(x) \\ &\le |s'-s| \int_s^{s'} \|v_r(x)\|^2_{\mu_r(x)} dr \end{aligned}$$

we get that for fixed $t \in [0,1]$, the map $s \mapsto \mathbf{T}(t,s,\cdot) \in L^2_{\mu_t}$ is absolutely continuous.

It can be proved that the set of regular curves is dense in the set of absolutely continuous curves on $\mathscr{P}_2(\mathbb{R}^d)$ with respect to uniform convergence plus convergence of length. We omit the technical proof of this fact and focus instead on the important case of geodesics:

Proposition 6.3 (Regular geodesics) *Let (μ_t) be a constant speed geodesic on $[0,1]$. Then its restriction to any interval $[\varepsilon, 1-\varepsilon]$, with $\varepsilon > 0$, is regular. In general, however, the whole curve (μ_t) may be not regular on $[0,1]$.*

Proof To prove that (μ_t) may be not regular just consider the case of $\mu_0 := \delta_x$ and $\mu_1 := \frac{1}{2}(\delta_{y_1} + \delta_{y_2})$: it is immediate to verify that for the velocity vector field (v_t) it holds $\mathrm{Lip}(v_t) = t^{-1}$.

For the other part, recall from Remark 2.25 (see also Proposition 2.16) that for $t \in (0,1)$ and $s \in [0,1]$ there exists a unique optimal map T_t^s from μ_t to μ_s. It is immediate to verify from formula (2.11) that these maps satisfy

$$\frac{T_t^s - Id}{s-t} = \frac{T_t^{s'} - Id}{s'-t}, \qquad \forall t \in (0,1),\ s \in [0,1].$$

Thus, thanks to Proposition 2.32, we have that v_t is given by

$$v_t = \lim_{s \to t} \frac{T_t^s - Id}{s-t} = \frac{Id - T_t^0}{t}. \tag{6.5}$$

Now recall that Remark 2.25 gives $\mathrm{Lip}(T_0^t) \leq (1-t)^{-1}$ to obtain

$$\mathrm{Lip}(v_t) \leq t^{-1}((1-t)^{-1} + 1) = \frac{2-t}{t(1-t)}.$$

Thus $t \mapsto \mathrm{Lip}(v_t)$ is integrable on any interval of the kind $[\varepsilon, 1-\varepsilon]$, $\varepsilon > 0$. □

Definition 6.4 (Vector fields along a curve) *A vector field along a curve (μ_t) is a Borel map $(t,x) \mapsto u_t(x)$ such that $u_t \in L^2_{\mu_t}$ for a.e. t. It will be denoted by (u_t).*

Observe that we are considering also non tangent vector fields, that is, we are not requiring $u_t \in \mathrm{Tan}_{\mu_t}(\mathscr{P}_2(\mathbb{R}^d))$ for a.e. t.

To define the (time) smoothness of a vector field (u_t) defined along a regular curve (μ_t) we will make an essential use of the flow maps: notice that the main problem in considering the smoothness of (u_t) is that for different times, the vectors belong to different spaces. To overcome this obstruction we will define the smoothness of $t \mapsto u_t \in L^2_{\mu_t}$ in terms of the smoothness of $t \mapsto u_t \circ \mathbf{T}(t_0,t,\cdot) \in L^2_{\mu_{t_0}}$:

Definition 6.5 (Smoothness of vector fields) *Let (μ_t) be a regular curve, $\mathbf{T}(t,s,\cdot)$ its flow maps and (u_t) a vector field defined along it. We say that (u_t) is absolutely continuous (or C^1, or C^n, …, or C^∞ or analytic) provided the map*

$$t \mapsto u_t \circ \mathbf{T}(t_0,t,\cdot) \in L^2_{\mu_{t_0}}$$

is absolutely continuous (or C^1, or C^n, …, or C^∞ or analytic) for every $t_0 \in [0,1]$.

Since $u_t \circ \mathbf{T}(t_1,t,\cdot) = u_t \circ \mathbf{T}(t_0,t,\cdot) \circ \mathbf{T}(t_1,t_0,\cdot)$ and the composition with $\mathbf{T}(t_1,t_0,\cdot)$ provides an isometry from $L^2_{\mu_{t_0}}$ to $L^2_{\mu_{t_1}}$, it is sufficient to check the regularity of $t \mapsto u_t \circ \mathbf{T}(t_0,t,\cdot)$ for *some* $t_0 \in [0,1]$ to be sure that the same regularity holds for every t_0.

Definition 6.6 (Total derivative) *With the same notation as above, assume that (u_t) is an absolutely continuous vector field. Its total derivative is defined as:*

$$\frac{\boldsymbol{d}}{dt}u_t := \lim_{h\to 0}\frac{u_{t+h}\circ \mathbf{T}(t,t+h,\cdot)-u_t}{h},$$

where the limit is intended in $L^2_{\mu_t}$.

Observe that we are not requiring the vector field to be tangent, and that the total derivative is in general a non tangent vector field, even if (u_t) is.

The identity

$$\begin{aligned}\lim_{h\to 0}\frac{u_{t+h}\circ \mathbf{T}(t,t+h,\cdot)-u_t}{h} &= \left(\lim_{h\to 0}\frac{u_{t+h}\circ \mathbf{T}(t_0,t+h,\cdot)-u_t\circ \mathbf{T}(t_0,t,\cdot)}{h}\right)\circ \mathbf{T}(t,t_0,\cdot)\\ &= \left(\frac{d}{dt}\big(u_t\circ \mathbf{T}(t_0,t,\cdot)\big)\right)\circ \mathbf{T}(t,t_0,\cdot),\end{aligned}$$

shows that the total derivative is well defined for a.e. t and that is an L^1 vector field, in the sense that it holds

$$\int_0^1 \left\|\frac{\boldsymbol{d}}{dt}u_t\right\|_{\mu_t} dt < \infty.$$

Notice also the inequality

$$\|u_s\circ \mathbf{T}(t,s,\cdot)-u_t\|_{\mu_t} \le \int_t^s \left\|\frac{d}{dt}(u_r\circ \mathbf{T}(t,r,\cdot))\right\|_{\mu_t} dr = \int_t^s \left\|\frac{\boldsymbol{d}}{dt}u_r\right\|_{\mu_r} dr.$$

An important property of the total derivative is the *Leibnitz rule*: for any couple of absolutely continuous vector fields (u^1_t), (u^2_t) along the same regular curve (μ_t) the map $t\mapsto \left\langle u^1_t,u^2_t\right\rangle_{\mu_t}$ is absolutely continuous and it holds

$$\frac{d}{dt}\left\langle u^1_t,u^2_t\right\rangle_{\mu_t} = \left\langle \frac{\boldsymbol{d}}{dt}u^1_t,u^2_t\right\rangle_{\mu_t} + \left\langle u^1_t,\frac{\boldsymbol{d}}{dt}u^2_t\right\rangle_{\mu_t}, \qquad a.e.\ t. \tag{6.6}$$

Indeed, from the identity

$$\left\langle u^1_t,u^2_t\right\rangle_{\mu_t} = \left\langle u^1_t\circ \mathbf{T}(t_0,t,\cdot),u^2_t\circ \mathbf{T}(t_0,t,\cdot)\right\rangle_{\mu_{t_0}},$$

it follows the absolute continuity, and the same expression gives

$$\begin{aligned}\frac{d}{dt}\left\langle u^1_t,u^2_t\right\rangle_{\mu_t} &= \frac{d}{dt}\left\langle u^1_t\circ \mathbf{T}(t_0,t,\cdot),u^2_t\circ \mathbf{T}(t_0,t,\cdot)\right\rangle_{\mu_{t_0}}\\ &= \left\langle \frac{d}{dt}\big(u^1_t\circ \mathbf{T}(t_0,t,\cdot)\big),u^2_t\circ \mathbf{T}(t_0,t,\cdot)\right\rangle_{\mu_{t_0}} + \left\langle u^1_t\circ \mathbf{T}(t_0,t,\cdot),\frac{d}{dt}\big(u^2_t\circ \mathbf{T}(t_0,t,\cdot)\big)\right\rangle_{\mu_{t_0}}\\ &= \left\langle \frac{\boldsymbol{d}}{dt}u^1_t,u^2_t\right\rangle_{\mu_t} + \left\langle u^1_t,\frac{\boldsymbol{d}}{dt}u^2_t\right\rangle_{\mu_t}.\end{aligned}$$

Example 6.7 (The smooth case) Let $(x,t)\mapsto \xi_t(x)$ be a C^∞_c vector field on $\mathbb{R}^d$, (μ_t) a regular curve and (v_t) its velocity vector field. Then the inequality

$$\|\xi_s\circ \mathbf{T}(t,s,\cdot)-\xi_t\|_{\mu_t} \le \|\xi_s-\xi_t\|_{\mu_s} + \|\xi_t\circ \mathbf{T}(t,s,\cdot)-\xi_t\|_{\mu_t} \le C|s-t| + C'\|\mathbf{T}(t,s,\cdot)-Id\|_{\mu_t},$$

with $C := \sup_{t,x} |\partial_t \xi_t(x)|$, $C' := \sup_{t,x} |\xi_t(x)|$, together with the fact that $s \mapsto \mathbf{T}(t, s, \cdot) \in L^2(\mu_t)$ is absolutely continuous, gives that (ξ_t) is absolutely continuous along (μ_t).

Then a direct application of the definition gives that its total derivative is given by

$$\frac{\boldsymbol{d}}{dt}\xi_t = \partial_t \xi_t + \nabla \xi_t \cdot v_t, \qquad a.e.\ t, \tag{6.7}$$

which shows that the total derivative is nothing but the *convective derivative* well known in fluid dynamics. ■

For $\mu \in \mathscr{P}_2(\mathbb{R}^d)$, we denote by $\mathrm{P}_\mu : L^2_\mu \to \mathrm{Tan}_\mu(\mathscr{P}_2(\mathbb{R}^d))$ the orthogonal projection, and we put $\mathrm{P}^\perp_\mu := Id - \mathrm{P}_\mu$.

Definition 6.8 (Covariant derivative) *Let (u_t) be an absolutely continuous and* tangent *vector field along the regular curve (μ_t). Its covariant derivative is defined as*

$$\frac{\boldsymbol{D}}{dt}u_t := \mathrm{P}_{\mu_t}\left(\frac{\boldsymbol{d}}{dt}u_t\right). \tag{6.8}$$

The trivial inequality

$$\left\|\frac{\boldsymbol{D}}{dt}u_t\right\|_{\mu_t} \leq \left\|\frac{\boldsymbol{d}}{dt}u_t\right\|_{\mu_t}$$

shows that the covariant derivative is an L^1 vector field.

In order to prove that the covariant derivative we just defined is the Levi-Civita connection, we need to prove two facts: *compatibiliy with the metric* and *torsion free identity*. Recall that on a standard Riemannian manifold, these two conditions are respectively given by:

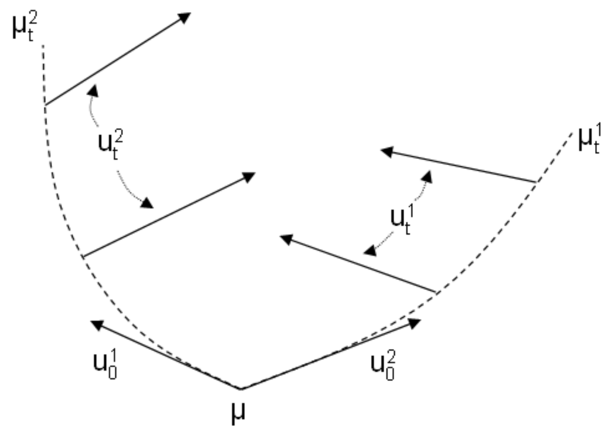
$$\begin{aligned}\frac{d}{dt}\langle X(\gamma_t), Y(\gamma_t)\rangle &= \left\langle (\nabla_{\gamma'_t}X)(\gamma_t), Y(\gamma_t)\right\rangle + \left\langle X(\gamma_t), (\nabla_{\gamma'_t}Y)(\gamma_t)\right\rangle \\ [X, Y] &= \nabla_X Y - \nabla_Y X,\end{aligned}$$

where X, Y are smooth vector fields and γ is a smooth curve on M.

The compatibility with the metric follows immediately from the Leibnitz rule (6.6), indeed if (u^1_t), (u^2_t) are tangent absolutely continuous vector fields we have:

$$\begin{aligned}\frac{d}{dt}\left\langle u^1_t, u^2_t\right\rangle_{\mu_t} &= \left\langle \frac{\boldsymbol{d}}{dt}u^1_t, u^2_t\right\rangle_{\mu_t} + \left\langle u^1_t, \frac{\boldsymbol{d}}{dt}u^2_t\right\rangle_{\mu_t} \\ &= \left\langle \mathrm{P}_{\mu_t}\left(\frac{\boldsymbol{d}}{dt}u^1_t\right), u^2_t\right\rangle_{\mu_t} + \left\langle u^1_t, \mathrm{P}_{\mu_t}\left(\frac{\boldsymbol{d}}{dt}u^2_t\right)\right\rangle_{\mu_t} \\ &= \left\langle \frac{\boldsymbol{D}}{dt}u^1_t, u^2_t\right\rangle_{\mu_t} + \left\langle u^1_t, \frac{\boldsymbol{D}}{dt}u^2_t\right\rangle_{\mu_t}.\end{aligned} \tag{6.9}$$

To prove the torsion-free identity, we need first to understand how to calculate the Lie bracket of two vector fields. To this aim, let μ^i_t, $i = 1,\ 2$, be two regular curves such that $\mu^1_0 = \mu^2_0 =: \mu$ and let $u^i_t \in \mathrm{Tan}_{\mu^i_t}(\mathscr{P}_2(\mathbb{R}^d))$ be two C^1 vector fields satisfying $u^1_0 = v^2_0$, $u^2_0 = v^1_0$, where v^i_t are the velocity vector fields of μ^i_t. We assume that the velocity fields v^i_t of μ^i_t are continuous in time (in the sense that the map $t \mapsto v^i_t \mu^i_t$ is continuous in the set of vector valued measure with the weak topology and $t \mapsto \|v^i_t\|_{\mu^i_t}$ is continuous as well), to be sure that (6.7) holds for *all* t with $v_t = v^i_t$ and the initial condition makes sense. With these hypotheses, it makes sense to consider the covariant derivative $\frac{\boldsymbol{D}}{dt}u^2_t$ along (μ^2_t) at $t = 0$: for this derivative we write $\boldsymbol{\nabla}_{u^1_0}u^2_t$. Similarly for (u^1_t).



Let us consider vector fields as derivations, and the functional $\mu \mapsto F_\varphi(\mu) := \int \varphi d\mu$, for given $\varphi \in C_c^\infty(\mathbb{R}^d)$. By the continuity equation, the derivative of F_φ along u_t^2 is equal to $\langle \nabla\varphi, u_t^2 \rangle_{\mu_t^2}$, therefore the compatibility with the metric (6.9) gives:

$$\begin{aligned} u^1(u^2(F_\varphi))(\mu) = \frac{d}{dt} \langle \nabla\varphi, u_t^2 \rangle_{\mu_t^2} |_{t=0} &= \langle \nabla^2\varphi \cdot v_0^2, u_0^2 \rangle_\mu + \left\langle \nabla\varphi, \boldsymbol{\nabla}_{u_0^1} u_t^2 \right\rangle_\mu \\ &= \langle \nabla^2\varphi \cdot u_0^1, u_0^2 \rangle_\mu + \left\langle \nabla\varphi, \boldsymbol{\nabla}_{u_0^1} u_t^2 \right\rangle_\mu . \end{aligned}$$

Subtracting the analogous term $u^2(u^1(F_\varphi))(\mu)$ and using the symmetry of $\nabla^2\varphi$ we get

$$[u^1, u^2](F_\varphi)(\mu) = \left\langle \nabla\varphi, \boldsymbol{\nabla}_{u_0^1} u_t^2 - \boldsymbol{\nabla}_{u_0^2} u_t^1 \right\rangle_\mu .$$

Given that the set $\{\nabla\varphi\}_{\varphi \in C_c^\infty}$ is dense in $\mathrm{Tan}_\mu(\mathscr{P}_2(\mathbb{R}^d))$, the above equation characterizes $[u^1, u^2]$ as:

$$[u^1, u^2] = \boldsymbol{\nabla}_{u_0^1} u_t^2 - \boldsymbol{\nabla}_{u_0^2} u_t^1, \tag{6.10}$$

which proves the torsion-free identity for the covariant derivative.

Example 6.9 (The velocity vector field of a geodesic) Let (μ_t) be the restriction to $[0,1]$ of a geodesic defined in some larger interval $(-\varepsilon, 1+\varepsilon)$ and let (v_t) be its velocity vector field. Then we know by Proposition 6.3 that (μ_t) is regular. Also, from formula (6.5) it is easy to see that it holds

$$v_s \circ \mathbf{T}(t, s, \cdot) = v_t, \qquad \forall t, s \in [0,1],$$

and thus (v_t) is absolutely continuous and satisfies $\frac{\boldsymbol{d}}{dt} v_t = 0$ and a fortiori $\frac{\boldsymbol{D}}{dt} v_t = 0$.

Thus, as expected, the velocity vector field of a geodesic has zero convariant derivative, in analogy with the standard Riemannian case. Actually, it is interesting to observe that not only the covariant derivative is 0 in this case, but also the total one. ■

Now we pass to the question of parallel transport. The definition comes naturally:

Definition 6.10 (Parallel transport) *Let (μ_t) be a regular curve. A tangent vector field (u_t) along it is a parallel transport if it is absolutely continuous and*

$$\frac{\boldsymbol{D}}{dt} u_t = 0, \qquad a.e.\ t.$$

It is immediate to verify that the scalar product of two parallel transports is preserved in time, indeed the compatibility with the metric (6.9) yields

$$\frac{d}{dt}\left\langle u_t^1, u_t^2 \right\rangle_{\mu_t} = \left\langle \frac{\boldsymbol{D}}{dt} u_t^1, u_t^2 \right\rangle_{\mu_t} + \left\langle u_t^1, \frac{\boldsymbol{D}}{dt} u_t^2 \right\rangle_{\mu_t} = 0, \quad a.e.\ t,$$

for any couple of parallel transports. In particular, this fact and the linearity of the notion of parallel transport give uniqueness of the parallel transport itself, in the sense that for any $u^0 \in \mathrm{Tan}_{\mu_0}(\mathscr{P}_2(\mathbb{R}^d))$ there exists at most one parallel transport (u_t) along (μ_t) satisfying $u_0 = u^0$.

Thus the problem is to show the existence. There is an important analogy, which helps understanding the proof, that we want to point out: we already know that the space $(\mathscr{P}_2(\mathbb{R}^d), W_2)$ looks like a Riemannian manifold, but actually it has also stronger similarities with a Riemannian manifold M embedded in some bigger space (say, on some Euclidean space $\mathbb{R}^D$), indeed in both cases:

- we have a natural presence of non tangent vectors: elements of $L^2_\mu \setminus \mathrm{Tan}_\mu(\mathscr{P}_2(\mathbb{R}^d))$ for $\mathscr{P}_2(\mathbb{R}^d)$, and vectors in $\mathbb{R}^D$ non tangent to the manifold for the embedded case.

- The scalar product in the tangent space can be naturally defined also for non tangent vectors: scalar product in L^2_μ for the space $\mathscr{P}_2(\mathbb{R}^d)$, and the scalar product in $\mathbb{R}^D$ for the embedded case. This means in particular that there are natural orthogonal projections from the set of tangent and non tangent vectors onto the set of tangent vectors: $\mathrm{P}_\mu : L^2_\mu \to \mathrm{Tan}_\mu(\mathscr{P}_2(\mathbb{R}^d))$ for $\mathscr{P}_2(\mathbb{R}^d)$ and $P_x : \mathbb{R}^D \to T_x M$ for the embedded case.

- The Covariant derivative of a tangent vector field is given by projecting the "time derivative" onto the tangent space. Indeed, for the space $\mathscr{P}_2(\mathbb{R}^d)$ we know that the covariant derivative is given by formula (6.8), while for the embedded manifold it holds:

$$\nabla_{\dot\gamma_t} u_t = P_{\gamma_t}\left(\frac{d}{dt} u_t\right), \tag{6.11}$$

where $t \mapsto \gamma_t$ is a smooth curve and $t \mapsto u_t \in T_{\gamma_t} M$ is a smooth tangent vector field.

Given these analogies, we are going to proceed as follows: first we give a proof of the existence of the parallel transport along a smooth curve in an embedded Riemannian manifold, then we will see how this proof can be adapted to the Wasserstein case: this approach should help highlighting what's the geometric idea behind the construction.

Thus, say that M is a given smooth Riemannian manifold embedded on $\mathbb{R}^D$, $t \mapsto \gamma_t \in M$ a smooth curve on $[0,1]$ and $u^0 \in T_{\gamma_0} M$ is a given tangent vector. Our goal is to prove the existence of an absolutely continuous vector field $t \mapsto u_t \in T_{\gamma_t} M$ such that $u_0 = u^0$ and

$$P_{\gamma_t}\left(\frac{d}{dt} u_t\right) = 0, \qquad a.e.\ t.$$

For any $t, s \in [0,1]$, let $\mathrm{tr}_t^s : T_{\gamma_t}\mathbb{R}^D \to T_{\gamma_s}\mathbb{R}^D$ be the natural translation map which takes a vector with base point γ_t (tangent or not to the manifold) and gives back the translated of this vector with base point γ_s. Notice that an effect of the curvature of the manifold and the chosen embedding on $\mathbb{R}^D$, is that $\mathrm{tr}_t^s(u)$ may be not tangent to M even if u is. Now define $P_t^s : T_{\gamma_t}\mathbb{R}^D \to T_{\gamma_s} M$ by

$$P_t^s(u) := P_{\gamma_s}(\mathrm{tr}_t^s(u)), \qquad \forall u \in T_{\gamma_t}\mathbb{R}^D.$$

An immediate consequence of the smoothness of M and γ are the two inequalities:

$$|\mathrm{tr}_t^s(u) - P_t^s(u)| \leq C|u||s-t|, \qquad \forall t,\ s \in [0,1] \text{ and } u \in T_{\gamma_t}M, \tag{6.12a}$$

$$|P_t^s(u)| \leq C|u||s-t|, \qquad \forall t,\ s \in [0,1] \text{ and } u \in T_{\gamma_t}^{\perp}M, \tag{6.12b}$$

where $T_{\gamma_t}^{\perp}M$ is the orthogonal complement of $T_{\gamma_t}M$ in $T_{\gamma_t}\mathbb{R}^D$. These two inequalities are all we need to prove existence of the parallel transport. The proof will be constructive, and is based on the identity:

$$\nabla_{\gamma_t} P_0^t(u)_{|t=0} = 0, \quad \forall u \in T_{\gamma(0)}M, \tag{6.13}$$

which tells that the vectors $P_0^t(u)$ are a first order approximation at $t=0$ of the parallel transport. Taking (6.11) into account, (6.13) is equivalent to

$$|P_t^0(\mathrm{tr}_0^t(u) - P_0^t(u))| = o(t), \qquad u \in T_{\gamma(0)}M. \tag{6.14}$$

Equation (6.14) follows by applying inequalities (6.12) (note that $\mathrm{tr}_0^t(u) - P_0^t(u) \in T_{\gamma_t}^{\perp}M$):

$$|P_t^0(\mathrm{tr}_0^t(u) - P_0^t(u))| \leq Ct|\mathrm{tr}_0^t(u) - P_0^t(u)| \leq C^2t^2|u|.$$

Now, let $\mathfrak{P}$ be the direct set of all the partitions of $[0,1]$, where, for $\mathcal{P},\ \mathcal{Q} \in \mathfrak{P}$, $\mathcal{P} \geq \mathcal{Q}$ if $\mathcal{P}$ is a refinement of $\mathcal{Q}$. For $\mathcal{P} = \{0 = t_0 < t_1 < \cdots < t_N = 1\} \in \mathfrak{P}$ and $u \in T_{\gamma_0}M$ define $\mathcal{P}(u) \in T_{\gamma_1}M$ as:

$$\mathcal{P}(u) := P_{t_{N-1}}^{t_N}(P_{t_{N-2}}^{t_{N-1}}(\cdots(P_0^{t_1}(u)))).$$

Our first goal is to prove that the limit $\mathcal{P}(u)$ for $\mathcal{P} \in \mathfrak{P}$ exists. This will naturally define a curve $t \to u_t \in T_{\gamma_t}M$ by taking partitions of $[0,t]$ instead of $[0,1]$: the final goal is to show that this curve is actually the parallel transport of u along the curve γ.

The proof is based on the following lemma.

Lemma 6.11 *Let $0 \leq s_1 \leq s_2 \leq s_3 \leq 1$ be given numbers. Then it holds:*

$$\left|P_{s_1}^{s_3}(u) - P_{s_2}^{s_3}(P_{s_1}^{s_2}(u))\right| \leq C^2|u||s_1 - s_2||s_2 - s_3|, \quad \forall u \in T_{\gamma_{s_1}}M.$$

Proof From $P_{s_1}^{s_3}(u) = P_{\gamma_{s_3}}(\mathrm{tr}_{s_1}^{s_3}(u)) = P_{\gamma_{s_3}}(\mathrm{tr}_{s_2}^{s_3}(\mathrm{tr}_{s_1}^{s_2}(u)))$ we get

$$P_{s_1}^{s_3}(u) - P_{s_2}^{s_3}(P_{s_1}^{s_2}(u)) = P_{s_2}^{s_3}(\mathrm{tr}_{s_1}^{s_2}(u) - P_{s_1}^{s_2}(u))$$

Since $u \in T_{\gamma_{s_1}}M$ and $\mathrm{tr}_{s_1}^{s_2}(u) - P_{s_1}^{s_2}(u) \in T_{\gamma_{s_2}}^{\perp}M$, the proof follows applying inequalities (6.12). $\square$

From this lemma, an easy induction shows that for any $0 \leq s_1 < \cdots < s_N \leq 1$ and $u \in T_{\gamma_{s_1}}M$ we have

$$\begin{aligned}
& \left|P_{s_1}^{s_N}(u) - P_{s_{N-1}}^{s_N}(P_{s_{N-2}}^{s_{N-1}}(\cdots(P_{s_1}^{s_2}(u))))\right| \\
\leq\ & \left|P_{s_1}^{s_N}(u) - P_{s_{N-1}}^{s_N}(P_{s_1}^{s_{N-1}}(u))\right| + \left|P_{s_{N-1}}^{s_N}(P_{s_1}^{s_{N-1}}(u)) - P_{s_{N-1}}^{s_N}(P_{s_{N-2}}^{s_{N-1}}(\cdots(P_{s_1}^{s_2}(u))))\right| \\
\leq\ & C^2|u||s_{N_1} - s_1|!s_N - s_{N-1}| + \left|P_{s_1}^{s_{N-1}}(u) - P_{s_{N-2}}^{s_{N-1}}(\cdots(P_{s_1}^{s_2}(u)))\right| \\
\leq\ & \cdots \\
\leq\ & C^2|u|\sum_{i=2}^{N-1}|s_1 - s_i||s_i - s_{i+1}| \leq C^2|u||s_1 - s_N|^2.
\end{aligned} \tag{6.15}$$

With this result, we can prove existence of the limit of $P(u)$ as P varies in $\mathfrak{P}$.

Theorem 6.12 *For any $u \in T_{\gamma_0}M$ there exists the limit of $\mathcal{P}(u)$ as $\mathcal{P}$ varies in $\mathfrak{P}$.*

Proof We have to prove that, given $\varepsilon > 0$, there exists a partition $\mathcal{P}$ such that

$$|\mathcal{P}(u) - \mathcal{Q}(u)| \leq |u|\varepsilon, \quad \forall \mathcal{Q} \geq \mathcal{P}. \tag{6.16}$$

In order to do so, it is sufficient to find $0 = t_0 < t_1 < \cdots < t_N = 1$ such that $\sum_i |t_{i+1} - t_i|^2 \leq \varepsilon/C^2$, and repeatedly apply equation (6.15) to all partitions induced by $\mathcal{Q}$ in the intervals (t_i, t_{i+1}). □

Now, for $s \leq t$ we can introduce the maps $T_t^s : T_{\gamma_t}M \to T_{\gamma_s}M$ which associate to the vector $u \in T_{\gamma_t}M$ the limit of the process just described taking into account partitions of $[s,t]$ instead of those of $[0,1]$.

Theorem 6.13 *For any $t_1 \leq t_2 \leq t_3 \in [0,1]$ it holds*

$$T_{t_2}^{t_3} \circ T_{t_1}^{t_2} = T_{t_1}^{t_3}. \tag{6.17}$$

Moreover, for any $u \in T_{\gamma_0}M$ the curve $t \to u_t := T_0^t(u) \in T_{\gamma_t}M$ is the parallel transport of u along γ.

Proof For the group property, consider those partitions of $[t_1, t_3]$ which contain t_2 and pass to the limit first on $[t_1, t_2]$ and then on $[t_2, t_3]$. To prove the second part of the statement, we prove first that (u_t) is absolutely continuous. To see this, pass to the limit in (6.15) with $s_1 = t_0$ and $s_N = t_1$, $u = u_{t_0}$ to get

$$|P_{t_0}^{t_1}(u_{t_0}) - u_{t_1}| \leq C^2|u_{t_0}|(t_1 - t_0)^2 \leq C^2|u|(t_1 - t_0)^2, \tag{6.18}$$

so that from (6.12a) we get

$$|\mathrm{tr}_{t_0}^{t_1}(u_{t_0}) - u_{t_1}| \leq |\mathrm{tr}_{t_0}^{t_1}(u_{t_0}) - P_{t_0}^{t_1}(u_{t_0})| + |P_{t_0}^{t_1}(u_{t_0}) - u_{t_1}| \leq C|u||t_1 - t_0|(1 + C|t_1 - t_0|),$$

which shows the absolute continuity. Finally, due to (6.17), it is sufficient to check that the covariant derivative vanishes at 0. To see this, put $t_0 = 0$ and $t_1 = t$ in (6.18) to get $|P_0^t(u) - u_t| \leq C^2|u|t^2$, so that the thesis follows from (6.13). □

Now we come back to the Wasserstein case. To follow the analogy with the Riemannian case, keep in mind that the analogous of the translation map tr_t^s is the right composition with $\mathbf{T}(s,t,\cdot)$, and the analogous of the map P_t^s is

$$\mathscr{P}_t^s(u) := \mathrm{P}_{\mu_s}(u \circ \mathbf{T}(s,t,\cdot)),$$

which maps $L^2_{\mu_t}$ onto $\mathrm{Tan}_{\mu_s}(\mathscr{P}_2(\mathbb{R}^d))$ We saw that the key to prove the existence of the parallel transport in the embedded Riemannian case are inequalities (6.12). Thus, given that we want to imitate the approach in the Wasserstein setting, we need to produce an analogous of those inequalities. This is the content of the following lemma.

We will denote by $\mathrm{Tan}_\mu^\perp(\mathscr{P}_2(\mathbb{R}^d))$ the orthogonal complement of $\mathrm{Tan}_\mu(\mathscr{P}_2(\mathbb{R}^d))$ in L^2_μ.

Lemma 6.14 (Control of the angles between tangent spaces) *Let $\mu, \nu \in \mathscr{P}_2(\mathbb{R}^d)$ and $T : \mathbb{R}^d \to \mathbb{R}^d$ be any Borel map satisfying $T_\# \mu = \nu$. Then it holds:*

$$\|v \circ T - \mathrm{P}_\mu(v \circ T)\|_\mu \leq \|v\|_\nu \mathrm{Lip}(T - Id), \qquad \forall v \in \mathrm{Tan}_\nu(\mathscr{P}_2(\mathbb{R}^d)),$$

and, if T is invertible, it also holds

$$\|\mathrm{P}_\mu(w \circ T)\|_\mu \leq \|w\|_\nu \mathrm{Lip}(T^{-1} - Id), \qquad \forall w \in \mathrm{Tan}_\nu^\perp(\mathscr{P}_2(\mathbb{R}^d)).$$

Proof We start with the first inequality, which is equivalent to

$$\|\nabla\varphi\circ T - P_\mu(\nabla\varphi\circ T)\|_\mu \le \|\nabla\varphi\|_\nu \mathrm{Lip}(T-Id), \qquad \forall\varphi\in C_c^\infty(\mathbb{R}^d). \tag{6.19}$$

Let us suppose first that $T-Id\in C_c^\infty(\mathbb{R}^d)$. In this case the map $\varphi\circ T$ is in $C_c^\infty(\mathbb{R}^d)$, too, and therefore $\nabla(\varphi\circ T)=\nabla T\cdot(\nabla\varphi)\circ T$ belongs to $\mathrm{Tan}_\mu(\mathscr{P}_2(\mathbb{R}^d))$. From the minimality properties of the projection we get:

$$\begin{aligned}\|\nabla\varphi\circ T - P_\mu(\nabla\varphi\circ T)\|_\mu &\le \|\nabla\varphi\circ T-\nabla T\cdot(\nabla\varphi)\circ T\|_\mu\\ &=\left(\int |(I-\nabla T(x))\cdot\nabla\varphi(T(x))|^2 d\mu(x)\right)^{1/2}\\ &\le\left(\int |\nabla\varphi(T(x))|^2\|\nabla(Id-T)(x)\|_{op}^2 d\mu(x)\right)^{1/2}\\ &\le \|\nabla\varphi\|_\nu \mathrm{Lip}(T-Id),\end{aligned}$$

where I is the identity matrix and $\|\nabla(Id-T)(x)\|_{op}$ is the operator norm of the linear functional from $\mathbb{R}^d$ to $\mathbb{R}^d$ given by $v\mapsto\nabla(Id-T)(x)\cdot v$.

Now turn to the general case, and we can certainly assume that T is Lipschitz. Then, it is not hard to see that there exists a sequence $(T_n-Id)\subset C_c^\infty(\mathbb{R}^d)$ such that $T_n\to T$ uniformly on compact sets and $\overline{\lim}_n \mathrm{Lip}(T_n-Id)\le\mathrm{Lip}(T-Id)$. It is clear that for such a sequence it holds $\|T-T_n\|_\mu\to 0$, and we have

$$\begin{aligned}\|\nabla\varphi\circ T - P_\mu(\nabla\varphi\circ T)\|_\mu &\le \|\nabla\varphi\circ T-\nabla(\varphi\circ T_n)\|_\mu\\ &\le\|\nabla\varphi\circ T-\nabla\varphi\circ T_n\|_\mu+\|\nabla\varphi\circ T_n-\nabla(\varphi\circ T_n)\|_\mu\\ &\le \mathrm{Lip}(\nabla\varphi)\|T-T_n\|_\mu+\|\nabla\varphi\circ T_n\|_\mu\mathrm{Lip}(T_n-Id).\end{aligned}$$

Letting $n\to+\infty$ we get the thesis.

For the second inequality, just notice that

$$\begin{aligned}\|\mathrm{P}_\mu(w\circ T)\|_\mu &= \sup_{\substack{v\in\mathrm{Tan}_\mu(\mathscr{P}_2(\mathbb{R}^d))\\ \|v\|_\mu=1}}\langle w\circ T, v\rangle_\mu = \sup_{\substack{v\in\mathrm{Tan}_\mu(\mathscr{P}_2(\mathbb{R}^d))\\ \|v\|_\mu=1}}\left\langle w, v\circ T^{-1}\right\rangle_\nu\\ &=\sup_{\substack{v\in\mathrm{Tan}_\mu(\mathscr{P}_2(\mathbb{R}^d))\\ \|v\|_\mu=1}}\left\langle w, v\circ T^{-1}-\mathrm{P}_\nu(v\circ T^{-1})\right\rangle_\nu\le\|w\|_\nu\mathrm{Lip}(T^{-1}-Id)\end{aligned}$$

□

From this lemma and the inequality

$$\mathrm{Lip}\Big(\mathbf{T}(s,t,\cdot)-Id\Big)\le e^{\left|\int_t^s\mathrm{Lip}(v_r)dr\right|}-1\le C\left|\int_t^s\mathrm{Lip}(v_r)dr\right|, \qquad \forall t,s\in[0,1],$$

(whose simple proof we omit), where $C:=e^{\int_0^1\mathrm{Lip}(v_r)dr}-1$, it is immediate to verify that it holds:

$$\begin{aligned}\|u\circ\mathbf{T}(s,t,\cdot)-\mathscr{P}_t^s(u)\|_{\mu_s}&\le C\|u\|_{\mu_t}\left|\int_t^s\mathrm{Lip}(v_r)dr\right|, \qquad u\in\mathrm{Tan}_{\mu_t}(\mathscr{P}_2(\mathbb{R}^d)),\\ \|\mathscr{P}_t^s(u)\|_{\mu_s}&\le C\|u\|_{\mu_t}\left|\int_t^s\mathrm{Lip}(v_r)dr\right|, \qquad u\in\mathrm{Tan}_{\mu_t}^\perp(\mathscr{P}_2(\mathbb{R}^d)).\end{aligned} \tag{6.20}$$

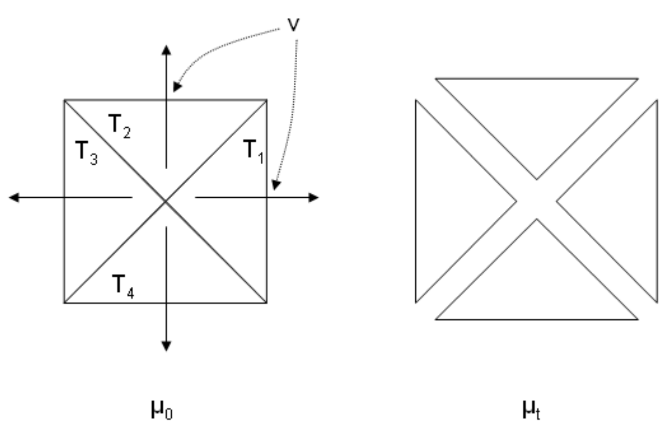
These inequalities are perfectly analogous to the (6.12) (well, the only difference is that here the bound on the angle is L^1 in t,s while for the embedded case it was L^∞, but this does not really change anything). Therefore the arguments presented before apply also to this case, and we can derive the existence of the parallel transport along regular curves:

Theorem 6.15 (Parallel transport along regular curves) *Let (μ_t) be a regular curve and $u^0 \in \mathrm{Tan}_{\mu_0}(\mathscr{P}_2(\mathbb{R}^d))$. Then there exists a parallel transport (u_t) along (μ_t) such that $u_0 = u^0$.*

Now, we know that the parallel transport exists along regular curves, and we know also that regular curves are dense, it is therefore natural to try to construct the parallel transport along any absolutely continuous curve via some limiting argument. However, this cannot be done, as the following counterexample shows:

Example 6.16 (Non existence of parallel transport along a non regular geodesic) Let $Q = [0,1] \times [0,1]$ be the unit square in $\mathbb{R}^2$ and let T_i, $i = 1,2,3,4$, be the four open triangles in which Q is divided by its diagonals. Let $\mu_0 := \chi_Q \mathscr{L}^2$ and define the function $v : Q \to \mathbb{R}^2$ as the gradient of the convex map $\max\{|x|,|y|\}$, as in the figure. Set also $w = v^\perp$, the rotation by $\pi/2$ of v, in Q and $w = 0$ out of Q. Notice that $\nabla \cdot (w\mu_0) = 0$.

Set $\mu_t := (Id + tv)_\# \mu_0$ and observe that, for positive t, the support Q_t of μ_t is made of 4 connected components, each one the translation of one of the sets T_i, and that $\mu_t = \chi_{Q_t} \mathscr{L}^2$.



It is immediate to check that (μ_t) is a geodesic in $[0,\infty)$, so that from 6.3 we know that the restriction of μ_t to any interval $[\varepsilon, 1]$ with $\varepsilon > 0$ is regular. Fix $\varepsilon > 0$ and note that, by construction, the flow maps of μ_t in $[\varepsilon, 1]$ are given by

$$\mathbf{T}(t,s,\cdot) = (Id + sv) \circ (Id + tv)^{-1}, \quad \forall t,s \in [\varepsilon, 1].$$

Now, set $w_t := w \circ \mathbf{T}(t,0,\cdot)$ and notice that w_t is tangent at μ_t (because w_t is constant in the connected components of the support of μ_t, so we can define a C_c^∞ function to be affine on each connected component and with gradient given by w_t, and then use the space between the components themselves to rearrange smoothly the function). Since $w_{t+h} \circ \mathbf{T}(t, t+h, \cdot) = w_t$, we have $\frac{\boldsymbol{d}}{dt} w_t = 0$ and a fortiori $\frac{\boldsymbol{D}}{dt} w_t = 0$. Thus (w_t) is a parallel transport in $[\varepsilon, 1]$. Furthermore, since $\nabla \cdot (w\mu_0) = 0$, we have $w_0 = w \notin \mathrm{Tan}_{\mu_0}(\mathscr{P}_2(\mathbb{R}^2))$. Therefore there is no way to extend w_t to a continuous *tangent* vector field on the whole $[0,1]$. In particular, there is no way to extend the parallel transport up to $t = 0$. ■

Now we pass to the calculus of total and covariant derivatives. Let (μ_t) be a fixed regular curve and let (v_t) be its velocity vector field. Start observing that, if (u_t) is absolutely continuous along

(μ_t), then $(\mathrm{P}_{\mu_t}(u_t))$ is absolutely continuous as well, as it follows from the inequality

$$\begin{aligned}
\Big\|\big(\mathrm{P}_{\mu_s}(u_s)\big)\circ\mathbf{T}(t,s,\cdot)-\mathrm{P}_{\mu_t}(u_t)\Big\|_{\mu_t} &\le \Big\|\big(\mathrm{P}_{\mu_s}(u_s)\big)\circ\mathbf{T}(t,s,\cdot)-\mathrm{P}_{\mu_t}\Big(\big(\mathrm{P}_{\mu_s}(u_s)\big)\circ\mathbf{T}(t,s,\cdot)\Big)\Big\|_{\mu_t}\\
&\quad+\Big\|\mathrm{P}_{\mu_t}\Big(\big(\mathrm{P}_{\mu_s}(u_s)\big)\circ\mathbf{T}(t,s,\cdot)\Big)-\mathrm{P}_{\mu_t}\big(u_s\circ\mathbf{T}(t,s,\cdot)\big)\Big\|_{\mu_t}\\
&\quad+\|\mathrm{P}_{\mu_t}(u_s\circ\mathbf{T}(t,s,\cdot))-\mathrm{P}_{\mu_t}(u_t)\|_{\mu_t}\\
&\le\Big\|\mathrm{P}^{\perp}_{\mu_t}\Big(\mathrm{P}_{\mu_s}(u_s)\circ\mathbf{T}(t,s,\cdot)\Big)\Big\|_{\mu_t}+\Big\|\mathrm{P}_{\mu_t}\Big(\mathrm{P}^{\perp}_{\mu_s}(u_s)\circ\mathbf{T}(t,s,\cdot)\Big)\Big\|_{\mu_t}\\
&\quad+\|u_s\circ\mathbf{T}(t,s,\cdot)-u_t\|_{\mu_t}\\
&\overset{(6.20)}{\le}2SC\int_t^s\mathrm{Lip}(v_r)dr+\int_t^s\left\|\frac{\boldsymbol{d}}{dr}u_r\right\|_{\mu_r}dr,
\end{aligned}\tag{6.21}$$

valid for any $t\le s$, where $S:=\sup_t\|u_t\|_{\mu_t}$. Thus $(\mathrm{P}_{\mu_t}(u_t))$ has a well defined covariant derivative for a.e. t. The question is: can we find a formula to express this derivative?

To compute it, apply the Leibniz rule for the total and covariant derivatives ((6.6) and (6.9)), to get that for a.e. $t\in[0,1]$ it holds

$$\begin{aligned}
\frac{d}{dt}\langle\mathrm{P}_{\mu_t}(u_t),\nabla\varphi\rangle_{\mu_t}&=\left\langle\frac{\boldsymbol{D}}{dt}\mathrm{P}_{\mu_t}(u_t),\nabla\varphi\right\rangle_{\mu_t}+\left\langle\mathrm{P}_{\mu_t}(u_t),\frac{\boldsymbol{D}}{dt}\nabla\varphi\right\rangle_{\mu_t},\\
\frac{d}{dt}\langle u_t,\nabla\varphi\rangle_{\mu_t}&=\left\langle\frac{\boldsymbol{d}}{dt}u_t,\nabla\varphi\right\rangle_{\mu_t}+\left\langle u_t,\frac{\boldsymbol{d}}{dt}\nabla\varphi\right\rangle_{\mu_t}.
\end{aligned}$$

Since $\nabla\varphi\in\mathrm{Tan}_{\mu_t}(\mathscr{P}_2(\mathbb{R}^d))$ for any t, it holds $\langle\mathrm{P}_{\mu_t}(u_t),\nabla\varphi\rangle_{\mu_t}=\langle u_t,\nabla\varphi\rangle_{\mu_t}$ for any $t\in[0,1]$, and thus the left hand sides of the previous equations are equal for a.e. t. Recalling formula (6.7) we have $\frac{\boldsymbol{d}}{dt}\nabla\varphi=\nabla^2\varphi\cdot v_t$ and $\frac{\boldsymbol{D}}{dt}\nabla\varphi=\mathrm{P}_{\mu_t}(\nabla^2\varphi\cdot v_t)$, thus from the equality of the right hand sides we obtain

$$\begin{aligned}
\left\langle\frac{\boldsymbol{D}}{dt}\mathrm{P}_{\mu_t}(u_t),\nabla\varphi\right\rangle_{\mu_t}&=\left\langle\frac{\boldsymbol{d}}{dt}u_t,\nabla\varphi\right\rangle_{\mu_t}+\langle u_t,\nabla^2\varphi\cdot v_t\rangle_{\mu_t}-\langle\mathrm{P}_{\mu_t}(u_t),\mathrm{P}_{\mu_t}(\nabla^2\varphi\cdot v_t)\rangle_{\mu_t}\\
&=\left\langle\frac{\boldsymbol{d}}{dt}u_t,\nabla\varphi\right\rangle_{\mu_t}+\langle\mathrm{P}^{\perp}_{\mu_t}(u_t),\mathrm{P}^{\perp}_{\mu_t}(\nabla^2\varphi\cdot v_t)\rangle_{\mu_t}.
\end{aligned}\tag{6.22}$$

This formula characterizes the scalar product of $\frac{\boldsymbol{D}}{dt}\mathrm{P}_{\mu_t}(u_t)$ with any $\nabla\varphi$ when φ varies on $C^\infty_c(\mathbb{R}^d)$. Since the set $\{\nabla\varphi\}$ is dense in $\mathrm{Tan}_{\mu_t}(\mathscr{P}_2(\mathbb{R}^d))$ for any $t\in[0,1]$, the formula actually identifies $\frac{\boldsymbol{D}}{dt}\mathrm{P}_{\mu_t}(u_t)$.

However, from this expression it is unclear what is the value of $\langle\frac{\boldsymbol{D}}{dt}\mathrm{P}_{\mu_t}(u_t),w\rangle_{\mu_t}$ for a general $w\in\mathrm{Tan}_{\mu_t}(\mathscr{P}_2(\mathbb{R}^d))$, because some regularity of $\nabla\varphi$ seems required to compute $\nabla^2\varphi\cdot v_t$. In order to better understand what the value of $\frac{\boldsymbol{D}}{dt}\mathrm{P}_{\mu_t}(u_t)$ is, fix $t\in[0,1]$ and assume for a moment that $v_t\in C^\infty_c(\mathbb{R}^d)$. Then compute the gradient of $x\mapsto\langle\nabla\varphi(x),v_t(x)\rangle$ to obtain

$$\nabla\langle\nabla\varphi,v_t\rangle=\nabla^2\varphi\cdot v_t+\nabla v_t^{\mathrm{t}}\cdot\nabla\varphi,$$

and consider this expression as an equality between vector fields in $L^2_{\mu_t}$. Taking the projection onto the Normal space we derive

$$\mathrm{P}^{\perp}_{\mu_t}(\nabla^2\varphi\cdot v_t)+\mathrm{P}^{\perp}_{\mu_t}(\nabla v_t^{\mathrm{t}}\cdot\nabla\varphi)=0.$$

Plugging the expression for $\mathrm{P}^\perp_{\mu_t}(\nabla^2\varphi\cdot v_t)$ into the formula for the covariant derivative we get

$$\begin{aligned}\left\langle \frac{\boldsymbol{D}}{dt}\mathrm{P}_{\mu_t}(u_t),\nabla\varphi\right\rangle_{\mu_t} &= \left\langle \frac{\boldsymbol{d}}{dt}u_t,\nabla\varphi\right\rangle_{\mu_t} - \left\langle \mathrm{P}^\perp_{\mu_t}(u_t),\mathrm{P}^\perp_{\mu_t}(\nabla v_t^{\mathrm{t}}\cdot\nabla\varphi)\right\rangle_{\mu_t}\\ &= \left\langle \frac{\boldsymbol{d}}{dt}u_t,\nabla\varphi\right\rangle_{\mu_t} - \left\langle \nabla v_t\cdot\mathrm{P}^\perp_{\mu_t}(u_t),\nabla\varphi\right\rangle_{\mu_t},\end{aligned}$$

which identifies $\frac{\boldsymbol{D}}{dt}\mathrm{P}_{\mu_t}(u_t)$ as

$$\frac{\boldsymbol{D}}{dt}\mathrm{P}_{\mu_t}(u_t) = \mathrm{P}_{\mu_t}\left(\frac{\boldsymbol{d}}{dt}u_t - \nabla v_t\cdot\mathrm{P}^\perp_{\mu_t}(u_t)\right). \tag{6.23}$$

We found this expression assuming that v_t was a smooth vector field, but given that we know that $\frac{\boldsymbol{D}}{dt}\mathrm{P}_{\mu_t}(u_t)$ exists for a.e. t, it is realistic to believe that the expression makes sense also for general Lipschitz v_t's. The problem is that the object ∇v_t may very well be not defined μ_t-a.e. for arbitrary μ_t and Lipschitz v_t (Rademacher's theorem is of no help here, because we are not assuming the measures μ_t to be absolutely continuous w.r.t. the Lebesgue measure). To give a meaning to formula (6.23) we need to introduce a new tensor.

Definition 6.17 (The Lipschitz non Lipschitz space) *Let $\mu\in\mathscr{P}_2(\mathbb{R}^d)$. The set $\mathrm{LNL}_\mu\subset[L^2_\mu]^2$ is the set of couples of vector fields (u,v) such that $\min\{\mathrm{Lip}(u),\mathrm{Lip}(v)\}<\infty$, i.e. the set of couples of vectors such that at least one of them is Lipschitz.*

We say that a sequence $(u_n,v_n)\in\mathrm{LNL}_\mu$ converges to $(u,v)\in\mathrm{LNL}_\mu$ provided $\|u_n-u\|_\mu\to0$, $\|v_n-v\|_\mu\to0$ and

$$\sup_n\min\{\mathrm{Lip}(u_n),\mathrm{Lip}(v_n)\}<\infty.$$

The following theorem holds:

Theorem 6.18 (The Normal tensor) *Let $\mu\in\mathscr{P}_2(\mathbb{R}^d)$. The map*

$$\begin{array}{rcl}\mathcal{N}_\mu(u,v):[C^\infty_c(\mathbb{R}^d,\mathbb{R}^d)]^2 & \to & \mathrm{Tan}^\perp_\mu(\mathscr{P}_2(\mathbb{R}^d)),\\ (u,v) & \mapsto & \mathrm{P}^\perp_\mu(\nabla u^{\mathrm{t}}\cdot v)\end{array}$$

extends uniquely to a sequentially continuous bilinear and antisymmetric map, still denoted by $\mathcal{N}_\mu$, from LNL_μ in $\mathrm{Tan}^\perp_\mu(\mathscr{P}_2(\mathbb{R}^d))$ for which the bound

$$\|\mathcal{N}_\mu(u,v)\|_\mu\le\min\{\mathrm{Lip}(u)\|v\|_\mu,\mathrm{Lip}(v)\|u\|_\mu\}, \tag{6.24}$$

holds.

Proof For $u,\ v\in C^\infty_c(\mathbb{R}^d,\mathbb{R}^d)$ we have $\nabla\langle u,v\rangle=\nabla u^{\mathrm{t}}\cdot v+\nabla v^{\mathrm{t}}\cdot u$ so that taking the projections on $\mathrm{Tan}^\perp_\mu(\mathscr{P}_2(\mathbb{R}^d))$ we get

$$\mathcal{N}_\mu(u,v)=-\mathcal{N}_\mu(v,u)\qquad\forall u,v\in C^\infty_c(\mathbb{R}^d,\mathbb{R}^d).$$

In this case, the bound (6.24) is trivial.

To prove existence and uniqueness of the sequentially continuous extension, it is enough to show that for any given sequence $n\mapsto(u_n,v_n)\in[C^\infty_c(\mathbb{R}^d,\mathbb{R}^d)]^2$ converging to some $(u,v)\in\mathrm{LNL}_\mu$, the sequence $n\mapsto\mathcal{N}_\mu(u_n,v_n)\in\mathrm{Tan}^\perp_\mu(\mathscr{P}_2(\mathbb{R}^d))$ is a Cauchy sequence. Fix such a sequence (u_n,v_n), let $L:=\sup_n\min\{\mathrm{Lip}(u_n),\mathrm{Lip}(v_n)\}$, $I\subset\mathbb{N}$ be the set of indexes n such that $\mathrm{Lip}(u_n)\le L$ and fix two smooth vectors $\tilde u,\tilde v\in C^\infty_c(\mathbb{R}^d,\mathbb{R}^d)$.

Notice that for $n, m \in I$ it holds

$$\begin{aligned}\|\mathcal{N}_\mu(u_n, v_n) - \mathcal{N}_\mu(u_m, v_m)\|_\mu &\le \|\mathcal{N}_\mu(u_n, v_n - \tilde{v})\|_\mu + \|\mathcal{N}_\mu(u_n - u_m, \tilde{v})\|_\mu + \|\mathcal{N}_\mu(u_m, \tilde{v} - v_m)\|_\mu \\ &\le L\|v_n - \tilde{v}\|_\mu + \mathrm{Lip}(\tilde{v})\|u_n - u_m\|_\mu + L\|v_m - \tilde{v}\|_\mu,\end{aligned}$$

and thus

$$\varlimsup_{\substack{n,m\to\infty \\ n,m\in I}} \|\mathcal{N}_\mu(u_n, v_n) - \mathcal{N}_\mu(u_m, v_m)\|_\mu \le 2L\|v - \tilde{v}\|_\mu,$$

(this expression being vacuum if I is finite). If $n \in I$ and $m \notin I$ we have $\mathrm{Lip}(v_m) \le L$ and

$$\begin{aligned}&\|\mathcal{N}_\mu(u_n, v_n) - \mathcal{N}_\mu(u_m, v_m)\|_\mu \\ &\quad\le \|\mathcal{N}_\mu(u_n, v_n - \tilde{v})\|_\mu + \|\mathcal{N}_\mu(u_n - \tilde{u}, \tilde{v})\|_\mu + \|\mathcal{N}_\mu(\tilde{u}, \tilde{v} - v_m)\|_\mu + \|\mathcal{N}_\mu(\tilde{u} - u_m, v_m)\|_\mu \\ &\quad\le L\|v_n - \tilde{v}\|_\mu + \mathrm{Lip}(\tilde{v})\|u_n - \tilde{u}\|_\mu + \mathrm{Lip}(\tilde{u})\|\tilde{v} - v_m\|_\mu + L\|u_m - \tilde{u}\|_\mu,\end{aligned}$$

which gives

$$\varlimsup_{\substack{n,m\to\infty \\ n\in I,\ m\notin I}} \|\mathcal{N}_\mu(u_n, v_n) - \mathcal{N}_\mu(u_m, v_m)\|_\mu \le L\|v - \tilde{v}\|_\mu + L\|u - \tilde{u}\|_\mu.$$

Exchanging the roles of the u's and the v's in these inequalities for the case in which $n \notin I$ we can conclude

$$\varlimsup_{n,m\to\infty} \|\mathcal{N}_\mu(u_n, v_n) - \mathcal{N}_\mu(u_m, v_m)\|_\mu \le 2L\|v - \tilde{v}\|_\mu + 2L\|u - \tilde{u}\|_\mu.$$

Since $\tilde{u}, \tilde{v}$ are arbitrary, we can let $\tilde{u} \to u$ and $\tilde{v} \to v$ in L^2_μ and conclude that $n \mapsto \mathcal{N}_\mu(u_n, v_n)$ is a Cauchy sequence, as requested.

The other claims follow trivially by the sequential continuity. □

Definition 6.19 (The operators $\mathcal{O}_v(\cdot)$ and $\mathcal{O}^*_v(\cdot)$) *Let $\mu \in \mathscr{P}_2(\mathbb{R})^d$ and $v \in L^2_\mu$ with $\mathrm{Lip}(v) < \infty$. Then the operator $u \mapsto \mathcal{O}_v(u)$ is defined by*

$$\mathcal{O}_v(u) := \mathcal{N}_\mu(v, u).$$

The operator $u \mapsto \mathcal{O}^*_v(u)$ is the adjoint of $\mathcal{O}_v(\cdot)$, i.e. it is defined by

$$\langle \mathcal{O}^*_v(u), w\rangle_\mu = \langle u, \mathcal{O}_v(w)\rangle_\mu, \qquad \forall w \in L^2_\mu.$$

It is clear that the operator norm of $\mathcal{O}_v(\cdot)$ and $\mathcal{O}^*_v(\cdot)$ is bounded by $\mathrm{Lip}(v)$. Observe that in writing $\mathcal{O}_v(u)$, $\mathcal{O}^*_v(u)$ we are losing the reference to the base measure μ, which certainly plays a role in the definition; this simplifies the notation and hopefully should create no confusion, as the measure we are referring to should always be clear from the context. Notice that if $v \in C^\infty_c(\mathbb{R}^d, \mathbb{R}^d)$ these operators read as

$$\begin{aligned}\mathcal{O}_v(u) &= \mathrm{P}^\perp_\mu(\nabla v^{\mathrm{t}} \cdot u), \\ \mathcal{O}^*_v(u) &= \nabla v \cdot \mathrm{P}^\perp_\mu(u).\end{aligned}$$

The introduction of the operators $\mathcal{O}_v(\cdot)$ and $\mathcal{O}^*_v(\cdot)$ allows to give a precise meaning to formula (6.23) for general regular curves:

Theorem 6.20 (Covariant derivative of $\mathrm{P}_{\mu_t}(u_t)$) *Let (μ_t) be a regular curve, (v_t) its velocity vector field and let (u_t) be an absolutely continuous vector field along it. Then $(\mathrm{P}_{\mu_t}(u_t))$ is absolutely continuous as well and for a.e. t it holds*

$$\frac{\boldsymbol{D}}{dt}\mathrm{P}_{\mu_t}(u_t) = \mathrm{P}_{\mu_t}\left(\frac{\boldsymbol{d}}{dt}u_t - \mathcal{O}^*_{v_t}(u_t)\right). \tag{6.25}$$

Proof The fact that $(\mathrm{P}_{\mu_t}(u_t))$ is absolutely continuous has been proved with inequality (6.21). To get the thesis, start from equation (6.22) and conclude noticing that for a.e. t it holds $\mathrm{Lip}(v_t) < \infty$ and thus

$$\mathrm{P}^{\perp}_{\mu_t}(\nabla^2\varphi \cdot v_t) = \mathcal{N}_\mu(\nabla\varphi, v_t) = -\mathcal{N}_\mu(v_t, \nabla\varphi) = -\mathcal{O}_{v_t}(\nabla\varphi)\,.$$

□

Corollary 6.21 (Total derivatives of $\mathrm{P}_{\mu_t}(u_t)$ and $\mathrm{P}^{\perp}_{\mu_t}(u_t)$) *Let (μ_t) be a regular curve, let (v_t) be its velocity vector field and let (u_t) be an absolutely continuous vector field along it. Then $(\mathrm{P}^{\perp}_{\mu_t}(u_t))$ is absolutely continuous and it holds*

$$\begin{aligned}
\frac{\boldsymbol{d}}{dt}\mathrm{P}_{\mu_t}(u_t) &= \mathrm{P}_{\mu_t}\left(\frac{\boldsymbol{d}}{dt}u_t\right) - \mathrm{P}_{\mu_t}\left(\mathcal{O}^*_{v_t}(u_t)\right) - \mathcal{O}_{v_t}\left(\mathrm{P}_{\mu_t}(u_t)\right),\\
\frac{\boldsymbol{d}}{dt}\mathrm{P}^{\perp}_{\mu_t}(u_t) &= \mathrm{P}^{\perp}_{\mu_t}\left(\frac{\boldsymbol{d}}{dt}u_t\right) + \mathrm{P}_{\mu_t}\left(\mathcal{O}^*_{v_t}(u_t)\right) + \mathcal{O}_{v_t}\left(\mathrm{P}_{\mu_t}(u_t)\right).
\end{aligned} \tag{6.26}$$

Proof The absolute continuity of $(\mathrm{P}^{\perp}_{\mu_t}(u_t))$ follows from the fact that both (u_t) and $(\mathrm{P}_{\mu_t}(u_t))$ are absolutely continuous. Similarly, the second formula in (6.26) follows immediately from the first one noticing that $u_t = \mathrm{P}_{\mu_t}(u_t) + \mathrm{P}^{\perp}_{\mu_t}(u_t)$ yields $\frac{\boldsymbol{d}}{dt}u_t = \frac{\boldsymbol{d}}{dt}\mathrm{P}_{\mu_t}(u_t) + \frac{\boldsymbol{d}}{dt}\mathrm{P}^{\perp}_{\mu_t}(u_t)$. Thus we have only to prove the first equality in (6.26). To this aim, let (w_t) be an arbitrary absolutely continuous vector field along (μ_t) and observe that it holds

$$\begin{aligned}
\frac{d}{dt}\left\langle \mathrm{P}_{\mu_t}(u_t), w_t\right\rangle_{\mu_t} &= \left\langle \frac{\boldsymbol{d}}{dt}\mathrm{P}_{\mu_t}(u_t), w_t\right\rangle_{\mu_t} + \left\langle \mathrm{P}_{\mu_t}(u_t), \frac{\boldsymbol{d}}{dt}w_t\right\rangle_{\mu_t},\\
\frac{d}{dt}\left\langle \mathrm{P}_{\mu_t}(u_t), \mathrm{P}_{\mu_t}(w_t)\right\rangle_{\mu_t} &= \left\langle \frac{\boldsymbol{D}}{dt}\mathrm{P}_{\mu_t}(u_t), \mathrm{P}_{\mu_t}(w_t)\right\rangle_{\mu_t} + \left\langle \mathrm{P}_{\mu_t}(u_t), \frac{\boldsymbol{D}}{dt}\mathrm{P}_{\mu_t}(w_t)\right\rangle_{\mu_t}.
\end{aligned}$$

Since the left hand sides of these expression are equal, the right hand sides are equal as well, thus we get

$$\begin{aligned}
\left\langle \frac{\boldsymbol{d}}{dt}\mathrm{P}_{\mu_t}(u_t) - \frac{\boldsymbol{D}}{dt}\mathrm{P}_{\mu_t}(u_t), w_t\right\rangle_{\mu_t} &= -\left\langle \mathrm{P}_{\mu_t}(u_t), \frac{\boldsymbol{d}}{dt}w_t - \frac{\boldsymbol{D}}{dt}\mathrm{P}_{\mu_t}(w_t)\right\rangle_{\mu_t}\\
&= -\left\langle \mathrm{P}_{\mu_t}(u_t), \mathrm{P}_{\mu_t}\Big(\frac{\boldsymbol{d}}{dt}w_t\Big) - \frac{\boldsymbol{D}}{dt}\mathrm{P}_{\mu_t}(w_t)\right\rangle_{\mu_t}\\
&\overset{(6.25)}{=} -\left\langle \mathrm{P}_{\mu_t}(u_t), \mathcal{O}^*_{v_t}(w_t)\right\rangle_{\mu_t}\\
&= -\left\langle \mathcal{O}_{v_t}\left(\mathrm{P}_{\mu_t}(u_t)\right), w_t\right\rangle_{\mu_t},
\end{aligned}$$

so that the arbitrariness of (w_t) gives

$$\frac{\boldsymbol{d}}{dt}\mathrm{P}_{\mu_t}(u_t) = \frac{\boldsymbol{D}}{dt}\mathrm{P}_{\mu_t}(u_t) - \mathcal{O}_{v_t}\left(\mathrm{P}_{\mu_t}(u_t)\right),$$

and the conclusion follows from (6.25). □

Along the same lines, the total derivative of $(\mathcal{N}_{\mu_t}(u_t, w_t))$ for given absolutely continuous vector fields (u_t), (w_t) along the same regular curve (μ_t) can be calculated. The only thing the we must take care of, is the fact that $\mathcal{N}_{\mu_t}$ is not defined on the whole $[L^2_{\mu_t}]^2$, so that we need to make some assumptions on (u_t), (w_t) to be sure that $(\mathcal{N}_{\mu_t}(u_t, w_t))$ is well defined and absolutely continuous. Indeed, observe that from a purely formal point of view, we expect that the total derivative of $(\mathcal{N}_{\mu_t}(u_t, w_t))$ is something like

$$\frac{\boldsymbol{d}}{dt}\mathcal{N}_{\mu_t}(u_t, w_t) = \mathcal{N}_{\mu_t}\left(\frac{\boldsymbol{d}}{dt}u_t, w_t\right)+\mathcal{N}_{\mu_t}\left(u_t, \frac{\boldsymbol{d}}{dt}w_t\right)+\left(\begin{array}{l}\text{some tensor - which we may think}\\ \text{as the derivative of } \mathcal{N}_{\mu_t} \text{ -}\\ \text{applied to the couple } (u_t, w_t)\end{array}\right).$$

Forget about the last object and look at the first two addends: given that the domain of definition of $\mathcal{N}_{\mu_t}$ is not the whole $[L^2_{\mu_t}]^2$, in order for the above formula to make sense, we should ask that in each of the couples $(\frac{\boldsymbol{d}}{dt}u_t, w_t)$ and $(u_t, \frac{\boldsymbol{d}}{dt}w_t)$, at least one vector is Lipschitz. Under the assumption that $\{\int_0^1 \mathrm{Lip}(u_t)dt < \infty \text{ and } \int_0^1 \mathrm{Lip}(\frac{\boldsymbol{d}}{dt}u_t)dt < +\infty\ \}$, it is possible to prove the following theorem (whose proof we omit).

Theorem 6.22 *Let (μ_t) be an absolutely continuous curve, let (v_t) be its velocity vector field and let (u_t), (w_t) be two absolutely continuous vector fields along it. Assume that $\int_0^1 \mathrm{Lip}(u_t)dt < \infty$ and $\int_0^1 \mathrm{Lip}(\frac{\boldsymbol{d}}{dt}u_t)dt < +\infty$. Then $(\mathcal{N}_{\mu_t}(u_t, w_t))$ is absolutely continuous and it holds*

$$\begin{aligned}\frac{\boldsymbol{d}}{dt}\mathcal{N}_{\mu_t}(u_t, w_t) =&\mathcal{N}_{\mu_t}\left(\frac{\boldsymbol{d}}{dt}u_t, w_t\right)+\mathcal{N}_{\mu_t}\left(u_t, \frac{\boldsymbol{d}}{dt}w_t\right)\\ &- \mathcal{O}_{v_t}\left(\mathcal{N}_{\mu_t}(u_t, w_t)\right) + \mathrm{P}_{\mu_t}\Big(\mathcal{O}^*_{v_t}\left(\mathcal{N}_{\mu_t}(u_t, w_t)\right)\Big).\end{aligned} \tag{6.27}$$

Corollary 6.23 *Let (μ_t) be a regular curve and assume that its velocity vector field (v_t) satisfies:*

$$\int_0^1 \mathrm{Lip}\left(\frac{\boldsymbol{d}}{dt}v_t\right)dt < \infty. \tag{6.28}$$

Then for every absolutely continuous vector field (u_t) both $(\mathcal{O}_{v_t}(u_t))$ and $(\mathcal{O}^*_{v_t}(u_t))$ are absolutely continuous and their total derivatives are given by:

$$\begin{aligned}\frac{\boldsymbol{d}}{dt}\mathcal{O}_{v_t}(u_t) &= \mathcal{O}_{\frac{\boldsymbol{d}}{dt}v_t}(u_t) + \mathcal{O}_{v_t}\left(\frac{\boldsymbol{d}}{dt}u_t\right) - \mathcal{O}_{v_t}\left(\mathcal{O}_{v_t}(u_t)\right) + \mathrm{P}_{\mu_t}\Big(\mathcal{O}^*_{v_t}\left(\mathcal{O}_{v_t}(u_t)\right)\Big)\\ \frac{\boldsymbol{d}}{dt}\mathcal{O}^*_{v_t}(u_t) &= \mathcal{O}^*_{\frac{\boldsymbol{d}}{dt}v_t}(u_t) + \mathcal{O}^*_{v_t}\left(\frac{\boldsymbol{d}}{dt}u_t\right) - \mathcal{O}^*_{v_t}\left(\mathcal{O}^*_{v_t}(u_t)\right) + \mathcal{O}^*_{v_t}\left(\mathcal{O}_{v_t}\left(\mathrm{P}_{\mu_t}(u_t)\right)\right)\end{aligned} \tag{6.29}$$

Proof The first formula follows directly from Theorem 6.22, the second from the fact that $\mathcal{O}^*_{v_t}(\cdot)$ is the adjoint of $\mathcal{O}_{v_t}(\cdot)$. $\square$

An important feature of equations (6.27) and (6.29) is that to express the derivatives of $(\mathcal{N}_{\mu_t}(u_t, w_t))$, $(\mathcal{O}_{v_t}(u_t))$ and $(\mathcal{O}^*_{v_t}(u_t))$ no "new operators appear". This implies that we can recursively calculate derivatives of any order of the vector fields $(\mathrm{P}_{\mu_t}(u_t))$, $(\mathrm{P}^\perp_{\mu_t}(u_t))$, $\mathcal{O}_{v_t}(u_t)$ and $\mathcal{O}^*_{v_t}(u_t)$, provided - of course - that we make appropriate regularity assumptions on the vector field (u_t) and on the velocity vector field (v_t). An example of result which can be proved following this direction is that the operator $t \mapsto \mathrm{P}_{\mu_t}(\cdot)$ is analytic along (the restriction of) a geodesic:

Proposition 6.24 (Analyticity of $t \mapsto \mathrm{P}_{\mu_t}(\cdot)$) *Let (μ_t) be the restriction to $[0, 1]$ of a geodesic defined in some larger interval $[-\varepsilon, 1 + \varepsilon]$. Then the operator $t \mapsto \mathrm{P}_{\mu_t}(\cdot)$ is analytic in the following sense. For any $t_0 \in [0, 1]$ there exists a sequence of bounded linear operators $A_n : L^2_{\mu_{t_0}} \to L^2_{\mu_{t_0}}$ such that the following equality holds in a neighborhood of t_0*

$$\mathrm{P}_{\mu_t}(u) = \sum_{n\in\mathbb{N}} \frac{(t-t_0)^n}{n!} A_n\big(u \circ \mathbf{T}(t_0, t, \cdot)\big) \circ \mathbf{T}(t, t_0, \cdot), \qquad \forall u \in L^2_{\mu_t}. \tag{6.30}$$

Proof From the fact that (μ_t) is the restriction of a geodesic we know that $L := \sup_{t\in[0,1]} \mathrm{Lip}(v_t) < \infty$ and that $\frac{\boldsymbol{d}}{dt}v_t = 0$ (recall Example 6.9). In particular condition (6.28) is fulfilled.

Fix $t_0 \in [0,1]$, $u \in L^2_{\mu_{t_0}}$ and define $u_t := u \circ \mathbf{T}(t, t_0, \cdot)$, so that $\frac{\boldsymbol{d}}{dt}u_t = 0$. From equations (6.26) and (6.29) and by induction it follows that $(\mathrm{P}_{\mu_t}(u_t))$ is C^∞. Also, $\frac{\boldsymbol{d}^n}{dt^n}\mathrm{P}_{\mu_t}(u_t)$ is the sum of addends each of which is the composition of projections onto the tangent or normal space and up to n operators $\mathcal{O}_{v_t}(\cdot)$ and $\mathcal{O}^*_{v_t}(\cdot)$, applied to the vector u_t. Since the operator norm of $\mathcal{O}_{v_t}(\cdot)$ and $\mathcal{O}^*_{v_t}(\cdot)$ is bounded by L, we deduce that

$$\left\| \frac{\boldsymbol{d}^n}{dt^n}\mathrm{P}_{\mu_t}(u_t) \right\|_{\mu_t} \leq \|u_t\|_{\mu_t} L^n = \|u\|_{\mu_{t_0}} L^n, \qquad \forall n \in \mathbb{N},\ t \in [0,1].$$

Defining the curve $t \mapsto U_t := \mathrm{P}_{\mu_t}(u_t) \circ \mathbf{T}(t_0, t, \cdot) \in L^2_{\mu_{t_0}}$, the above bound can be written as

$$\left\| \frac{d^n}{dt^n}U_t \right\|_{\mu_{t_0}} \leq \|U_{t_0}\|_{\mu_{t_0}} L^n, \qquad \forall n \in \mathbb{N},\ t \in [0,1],$$

which implies that the curve $t \mapsto U_t \in L^2_{\mu_{t_0}}$ is analytic. This means that for t close to t_0 it holds

$$\mathrm{P}_{\mu_t}(u_t) \circ \mathbf{T}(t_0, t, \cdot) = \sum_{n\geq 0} \frac{(t-t_0)^n}{n!} \frac{\boldsymbol{d}^n}{dt^n}|_{t=t_0}(\mathrm{P}_{\mu_t}(u_t)).$$

Now notice that equations (6.26) and (6.29) and the fact that $\frac{\boldsymbol{d}}{dt}u_t \equiv 0$ ensure that $\frac{\boldsymbol{d}^n}{dt^n}|_{t=t_0}(\mathrm{P}_{\mu_t}(u_t)) = A_n(u)$, where $A_n : L^2_{\mu_{t_0}} \to L^2_{\mu_{t_0}}$ is bounded. Thus the thesis follows by the arbitrariness of $u \in L^2_{\mu_{t_0}}$. $\square$

Now we have all the technical tools we need in order to study the curvature tensor of the "manifold" $\mathscr{P}_2(\mathbb{R}^d)$.

Following the analogy with the Riemannian case, we are lead to define the curvature tensor in the following way: given three vector fields $\mu \mapsto \nabla\varphi^i_\mu \in \mathrm{Tan}_\mu(\mathscr{P}_2(\mathbb{R}^d))$, $i = 1, \ldots, 3$, the curvature tensor $\mathbf{R}$ calculated on them at the measure μ is defined as:

$$\mathbf{R}(\nabla\varphi^1_\mu, \nabla\varphi^2_\mu)(\nabla\varphi^3_\mu) := \boldsymbol{\nabla}_{\nabla\varphi^2_\mu}(\boldsymbol{\nabla}_{\nabla\varphi^1_\mu}\nabla\varphi^3_\mu) - \boldsymbol{\nabla}_{\nabla\varphi^1_\mu}(\boldsymbol{\nabla}_{\nabla\varphi^2_\mu}\nabla\varphi^3_\mu) + \boldsymbol{\nabla}_{[\nabla\varphi^1_\mu, \nabla\varphi^2_\mu]}\nabla\varphi^3_\mu,$$

where the objects like $\boldsymbol{\nabla}_{\nabla\varphi_\mu}(\nabla\psi_\mu)$, are, heuristically speaking, the covariant derivative of the vector field $\mu \mapsto \nabla\psi_\mu$ along the vector field $\mu \mapsto \nabla\varphi_\mu$.

However, in order to give a precise meaning to the above formula, we should be sure, at least, that the derivatives we are taking exist. Such an approach is possible, but heavy: indeed, consider that we should define what are C^1 and C^2 vector fields, and in doing so we cannot just consider derivatives along curves. Indeed we would need to be sure that "the partial derivatives have the right symmetries", otherwise there won't be those cancellations which let the above operator be a tensor.

Instead, we adopt the following strategy:

- First we calculate the curvature tensor for some very specific kind of vector fields, for which we are able to do and justify the calculations. Specifically, we will consider vector fields of the kind $\mu \mapsto \nabla\varphi$, where the function $\varphi \in C^\infty_c(M)$ does not depend on the measure μ.
- Then we prove that the object found is actually a tensor, i.e. that its value depends only on the $\mu-$a.e. value of the considered vector fields, and not on the fact that we obtained the formula assuming that the functions φ's were independent on the measure.

- Finally, we discuss the minimal regularity requirements for the object found to be well defined.

Pick $\varphi, \psi \in C_c^\infty(\mathbb{R}^d)$ and observe that a curve of the kind $t \mapsto (Id + t\nabla\varphi)_\# \mu$ is a regular geodesic on an interval $[-T, T]$ for T sufficiently small (Remark 1.22 and Proposition 6.3). It is then immediate to verify that a vector field of the kind $(\nabla\psi)$ along it is C^∞. Its covariant derivative calculated at $t = 0$ is given by $\mathrm{P}_\mu(\nabla^2\psi \cdot \nabla\varphi)$. Thus we write:

$$\boldsymbol{\nabla}_{\nabla\varphi}\nabla\psi := \mathrm{P}_\mu(\nabla^2\psi \cdot \nabla\varphi) \qquad \forall \varphi, \psi \in C_c^\infty(\mathbb{R}^d). \tag{6.31}$$

Proposition 6.25 *Let $\mu \in \mathscr{P}_2(\mathbb{R}^d)$ and $\varphi_1, \varphi_2, \varphi_3 \in C_c^\infty(\mathbb{R}^d)$. The curvature tensor $\mathbf{R}$ in μ calculated for the 3 vector fields $\nabla\varphi_i$, $i = 1, 2, 3$ is given by*

$$\begin{aligned}\mathbf{R}(\nabla\varphi_1, \nabla\varphi_2)\nabla\varphi_3 = &\mathrm{P}_\mu\Big(\mathcal{O}^*_{\nabla\varphi_2}\left(\mathcal{N}_\mu(\nabla\varphi_1, \nabla\varphi_3)\right) \\ &- \mathcal{O}^*_{\nabla\varphi_1}\left(\mathcal{N}_\mu(\nabla\varphi_2, \nabla\varphi_3)\right) + 2\mathcal{O}^*_{\nabla\varphi_3}\left(\mathcal{N}_\mu(\nabla\varphi_1, \nabla\varphi_2)\right)\Big).\end{aligned} \tag{6.32}$$

Proof We start computing the value of $\boldsymbol{\nabla}_{\nabla\varphi_2}\boldsymbol{\nabla}_{\nabla\varphi_1}\nabla\varphi_3$. Let $\mu_t := (Id + t\nabla\varphi_2)_\#\mu$ and observe, as just recalled, that (μ_t) is a regular geodesic in some symmetric interval $[-T, T]$. The vector field $\nabla^2\varphi_3 \cdot \nabla\varphi_1$ is clearly C^∞ along it, thus by Proposition 6.24 also the vector field $u_t := \mathrm{P}_{\mu_t}(\nabla^2\varphi_3 \cdot \nabla\varphi_1) = \boldsymbol{\nabla}_{\nabla\varphi_1}\nabla\varphi_3(\mu_t)$ is C^∞. The covariant derivative at $t = 0$ of (u_t) along (μ_t) is, by definition, the value of $\boldsymbol{\nabla}_{\nabla\varphi_2}\boldsymbol{\nabla}_{\nabla\varphi_1}\nabla\varphi_3$ at μ. Applying formula (6.25) we get

$$\boldsymbol{\nabla}_{\nabla\varphi_2}\boldsymbol{\nabla}_{\nabla\varphi_1}\nabla\varphi_3 = \mathrm{P}_\mu\left(\nabla(\nabla^2\varphi_3 \cdot \nabla\varphi_1) \cdot \nabla\varphi_2 - \nabla^2\varphi_2 \cdot \mathrm{P}^\perp_\mu(\nabla^2\varphi_3 \cdot \nabla\varphi_1)\right). \tag{6.33}$$

Symmetrically, it holds

$$\boldsymbol{\nabla}_{\nabla\varphi_1}\boldsymbol{\nabla}_{\nabla\varphi_2}\nabla\varphi_3 = \mathrm{P}_\mu\left(\nabla(\nabla^2\varphi_3 \cdot \nabla\varphi_2) \cdot \nabla\varphi_1 - \nabla^2\varphi_1 \cdot \mathrm{P}^\perp_\mu(\nabla^2\varphi_3 \cdot \nabla\varphi_2)\right). \tag{6.34}$$

Finally, from the torsion free identity (6.10) we have

$$[\nabla\varphi_1, \nabla\varphi_2] = \mathrm{P}_\mu(\nabla^2\varphi_1 \cdot \nabla\varphi_2 - \nabla^2\varphi_2 \cdot \nabla\varphi_1),$$

and thus

$$\boldsymbol{\nabla}_{[\nabla\varphi_1, \nabla\varphi_2]}\nabla\varphi_3 = \mathrm{P}_\mu\Big(\nabla^2\varphi_3 \cdot \Big(\mathrm{P}_\mu(\nabla^2\varphi_1 \cdot \nabla\varphi_2 - \nabla^2\varphi_2 \cdot \nabla\varphi_1)\Big)\Big). \tag{6.35}$$

Subtracting (6.35) and (6.34) from (6.33) and observing that

$$\nabla(\nabla^2\varphi_3 \cdot \nabla\varphi_1) \cdot \nabla\varphi_2 - \nabla(\nabla^2\varphi_3 \cdot \nabla\varphi_2) \cdot \nabla\varphi_1 = \nabla^2\varphi_3 \cdot \nabla^2\varphi_1 \cdot \nabla\varphi_2 - \nabla^2\varphi_3 \cdot \nabla^2\varphi_2 \cdot \nabla\varphi_1,$$

we get the thesis. □

Observe that equation (6.32) is equivalent to

$$\begin{aligned}\langle\mathbf{R}(\nabla\varphi_1, \nabla\varphi_2)\nabla\varphi_3, \nabla\varphi_4\rangle_\mu = &\langle\mathcal{N}_\mu(\nabla\varphi_1, \nabla\varphi_3), \mathcal{N}_\mu(\nabla\varphi_2, \nabla\varphi_4)\rangle_\mu \\ &- \langle\mathcal{N}_\mu(\nabla\varphi_2, \nabla\varphi_3), \mathcal{N}_\mu(\nabla\varphi_1, \nabla\varphi_4)\rangle_\mu \\ &+ 2\,\langle\mathcal{N}_\mu(\nabla\varphi_1, \nabla\varphi_2), \mathcal{N}_\mu(\nabla\varphi_3, \nabla\varphi_4)\rangle_\mu,\end{aligned} \tag{6.36}$$

for any $\varphi_4 \in C_c^\infty(M)$. From this formula it follows immediately that the operator $\mathbf{R}$ is actually a tensor:

Proposition 6.26 *Let $\mu \in \mathscr{P}_2(\mathbb{R}^d)$. The curvature operator, given by formula* (6.36)*, is a tensor on $[\{\nabla\varphi\}]^4$, i.e. its value depends only on the $\mu-$a.e. value of the 4 vector fields.*

Proof Clearly the left hand side of equation (6.36) is a tensor w.r.t. the fourth entry. The conclusion follows from the symmetries of the right hand side. □

We remark that from (6.36) it follows that $\mathbf{R}$ has all the expected symmetries.

Concerning the domain of definition of the curvature tensor, the following statement holds, whose proof follows from the properties of the normal tensor $\mathcal{N}_\mu$:

Proposition 6.27 *Let $\mu \in \mathscr{P}_2(\mathbb{R}^d)$. Then the curvature tensor, thought as map from $[\{\nabla\varphi\}]^4$ to $\mathbb{R}$ given by* (6.36)*, extends uniquely to a sequentially continuous map on the set of 4-ples of vector fields in L^2_μ in which at least 3 vector fields are Lipschitz, where we say that $(v^1_n, v^2_n, v^3_n, v^4_n)$ is converging to (v^1, v^2, v^3, v^4) if there is convergence in L^2_μ on each coordinate and*

$$\sup_n \mathrm{Lip}(v^i_n) < \infty,$$

for at least 3 indexes i.

Thus, in order for the curvature tensor to be well defined we need at least 3 of the 4 vector fields involved to be Lipschitz. However, for some related notion of curvature the situation simplifies. Of particular relevance is the case of sectional curvature:

Example 6.28 (The sectional curvature) If we evaluate the curvature tensor $\mathbf{R}$ on a 4-ple of vectors of the kind (u, v, u, v) and we recall the antisymmetry of $\mathcal{N}_\mu$ we obtain

$$\langle \mathbf{R}(u, v)u, v\rangle_\mu = 3\,\|\mathcal{N}_\mu(u, v)\|^2_\mu .$$

Thanks to the simplification of the formula, the value of $\langle \mathbf{R}(u, v)u, v\rangle_\mu$ is well defined as soon as either u or v is Lipschitz. That is, $\langle \mathbf{R}(u, v)u, v\rangle_\mu$ is well defined for $(u, v) \in \mathrm{L}\textsc{n}\mathrm{L}_\mu$. In analogy with the Riemannian case we can therefore define the sectional curvature $\mathbf{K}(u, v)$ at the measure μ along the directions u, v by

$$\mathbf{K}(u, v) := \frac{\langle \mathbf{R}(u, v)u, v\rangle_\mu}{\|u\|^2_\mu\|v\|^2_\mu - \langle u, v\rangle^2_\mu} = \frac{3\,\|\mathcal{N}_\mu(u, v)\|^2_\mu}{\|u\|^2_\mu\|v\|^2_\mu - \langle u, v\rangle^2_\mu}, \qquad \forall (u, v) \in \mathrm{L}\textsc{n}\mathrm{L}_\mu.$$

This expression confirms the fact that the sectional curvatures of $\mathscr{P}_2(\mathbb{R}^d)$ are positive (coherently with Theorem 2.20), and provides a rigorous proof of the analogous formula already appeared in [67] and formally computed using O'Neill formula. ■

6.4 Bibliographical notes

The idea of looking at the Wasserstein space as a sort of infinite dimensional Riemannian manifold is due to F. Otto and given in his seminal paper [67]. The whole discussion in Section 6.1 is directly taken from there.

The fact that the 'tangent space made of gradients' $\mathrm{Tan}_\mu(\mathscr{P}_2(\mathbb{R}^d))$ was not sufficient to study all the aspects of the 'Riemannian geometry' of $(\mathscr{P}_2(\mathbb{R}^d), W_2)$ has been understood in [6] in connection with the definition of subdifferential of a geodesically convex functional, in particular concerning the issue of having a closed subdifferential. In the appendix of [6] the concept of Geometric Tangent space discussed in Section 6.2 has been introduced. Further studies on the properties of $\mathbf{Tan}_\mu(\mathscr{P}_2(M))$ have been made in [43]. Theorem 6.1 has been proved in [46].

The first work in which a description of the covariant derivative and the curvature tensor of $(\mathscr{P}_2(M), W_2)$, M being a compact Riemannian manifold has been given (beside the formal calculus of the sectional curvature via O'Neill formula done already in [67]) is the paper of J. Lott [56]: rigorous formulas are derived for the computation of such objects on the 'submanifold' $\mathscr{P}_{C^\infty}(M)$

made of absolutely continuous measures with density C^∞ and bounded away from 0. In the same paper Lott shows that if M has a Poisson structure, then the same is true for $\mathscr{P}_{C^\infty}(M)$ (a topic which has not been addressed in these notes).

Independently on Lott's work, the second author built the parallel transport on $(\mathscr{P}_2(\mathbb{R}^d), W_2)$ in his PhD thesis [43], along the same lines provided in Section 6.3. The differences with Lott's work are the fact that the analysis was carried out on $\mathbb{R}^d$ rather than on a compact Riemannian manifold, that no assumptions on the measures were given, and that both the existence Theorem 6.15 for the parallel transport along a regular curve and counterexamples to its general existence (the Example 6.16) were provided. These results have been published by the authors of these notes in [5]. Later on, after having beed aware of Lott's results, the second author generalized the construction to the case of Wasserstein space built over a manifold in [44]. Not all the results have been reported here: we mention that it is possible to push the analysis up show the differentiability properties of the exponential map and the existence of Jacobi fields.

7 Ricci curvature bounds

Let us start recalling what is the Ricci curvature for a Riemannian manifold M (which we will always consider smooth and complete). Let R be the Riemann curvature tensor on M, $x \in M$ and $u,\, v \in T_xM$. Then the Ricci curvature $\mathrm{Ric}(u,v) \in \mathbb{R}$ is defined as

$$\mathrm{Ric}(u,v) := \sum_i \langle R(u,e_i)v, e_i\rangle\,,$$

where $\{e_i\}$ is any orthonormal basis of T_xM. An immediate consequence of the definition and the symmetries of R is the fact that $\mathrm{Ric}(u,v) = \mathrm{Ric}(v,u)$.

Another, more geometric, characterization of the Ricci curvature is the following. Pick $x \in M$, a small ball B around the origin in T_xM and let μ be the Lebesgue measure on B. The exponential map $\exp_x : B \to M$ is injective and smooth, thus the measure $(\exp_x)_\#\mu$ has a smooth density w.r.t. the volume measure Vol on M. For any $u \in B$, let $f(u)$ be the density of $(\exp_x)_\#\mu$ w.r.t. Vol at the point $\exp_x(u)$. Then the function f has the following Taylor expansion:

$$f(u) = 1 + \frac{1}{2}\mathrm{Ric}(u,u) + o(|u|^2). \tag{7.1}$$

It is said that the Ricci curvature is bounded below by $\lambda \in \mathbb{R}$ provided

$$\mathrm{Ric}(u,u) \geq \lambda |u|^2,$$

for every $x \in M$ and $u \in T_xM$.

Several important geometric and analytic inequalities are related to bounds from below on Ricci curvature, we mention just two of them.

- Brunn-Minkowski. Suppose that M has non negative Ricci curvature, and for any $A_0, A_1 \subset M$ compact, let

  $$A_t := \Big\{\gamma_t \;:\; \gamma \text{ is a constant speed geodesic s.t. } \gamma_0 \in A_0,\ \gamma_1 \in A_1\Big\}, \qquad \forall t \in [0,1].$$

  Then it holds

  $$\big(\mathrm{Vol}(A_t)\big)^{1/n} \geq (1-t)\big(\mathrm{Vol}(A_0)\big)^{1/n} + t\big(\mathrm{Vol}(A_1)\big)^{1/n}, \qquad \forall t \in [0,1], \tag{7.2}$$

  where n is the dimension of M.

- Bishop-Gromov. Suppose that M has Ricci curvature bounded from below by $(n-1)k$, where n is the dimension of M and k a real number. Let $\tilde{M}$ be the simply connected, n-dimensional space with constant curvature, having Ricci curvature equal to $(n-1)k$ (so that $\tilde{M}$ is a sphere if $k > 0$, a Euclidean space if $k = 0$ and an hyperbolic space if $k < 0$). Then for every $x \in M$ and $\tilde{x} \in \tilde{M}$ the map
$$(0, \infty) \ni r \quad \mapsto \quad \frac{\mathrm{Vol}(B_r(x))}{\widetilde{\mathrm{Vol}}(B_r(\tilde{x}))}, \tag{7.3}$$
is non increasing, where Vol and $\widetilde{\mathrm{Vol}}$ are the volume measures on M, $\tilde{M}$ respectively.

A natural question is whether it is possible to formulate the notion of Ricci bound from below also for metric spaces, analogously to the definition of Alexandrov spaces, which are a metric analogous of Riemannian manifolds with bounded (either from above or from below) sectional curvature. What became clear over time, is that the correct non-smooth object where one could try to give a notion of Ricci curvature bound is not a metric space, but rather a metric *measure* space, i.e. a metric space where a reference non negative measure is also given. When looking to the Riemannian case, this fact is somehow hidden, as a natural reference measure is given by the volume measure, which is a function of the distance.

There are several viewpoints from which one can see the necessity of a reference measure (which can certainly be the Hausdorff measure of appropriate dimension, if available). A first (cheap) one is the fact that in most of identities/inequalities where the Ricci curvature appears, also the reference measures appears (e.g. equations (7.1), (7.2) and (7.3) above). A more subtle point of view comes from studying stability issues: consider a sequence (M_n, g_n) of Riemannian manifolds and assume that it converges to a smooth Riemannian manifold (M, g) in the Gromov-Hausdorff sense. Assume that the Ricci curvature of (M_n, g_n) is uniformly bounded below by some $K \in R$. Can we deduce that the Ricci curvature of (M, g) is bounded below by K? The answer is *no* (while the same question with sectional curvature in place of Ricci one has affirmative answer). It is possible to see that when Ricci bounds are not preserved in the limiting process, it happens that the volume measures of the approximating manifolds are not converging to the volume measure of the limit one.

Another important fact to keep in mind is the following: if we want to derive useful analytic/geometric consequences from a weak definition of Ricci curvature bound, we should also known what is the dimension of the metric measure space we are working with: consider for instance the Brunn-Minkowski and the Bishop-Gromov inequalities above, both make sense if we know the dimension of M, and not just that its Ricci curvature is bounded from below. This tells that the natural notion of bound on the Ricci curvature should be a notion speaking both about the *curvature* and the *dimension* of the space. Such a notion exists and is called $CD(K, N)$ condition, K being the bound from below on the Ricci curvature, and N the bound from above on the dimension. Let us tell in advance that we will focus only on two particular cases: the curvature dimension condition $CD(K, \infty)$, where no upper bound on the dimension is specified, and the curvature-dimension condition $CD(0, N)$, where the Ricci curvature is bounded below by 0. Indeed, the general case is much more complicated and there are still some delicate issues to solve before we can say that the theory is complete and fully satisfactory.

Before giving the definition, let us highlight which are the qualitative properties that we expect from a weak notion of curvature-dimension bound:

Intrinsicness. The definition is based only on the property of the space itself, that is, is not something like "if the space is the limit of smooth spaces...."

Compatibility. If the metric-measure space is a Riemannian manifold equipped with the volume measure, then the bound provided by the abstract definition coincides with the lower bound on the

Ricci curvature of the manifold, equipped with the Riemannian distance and the volume measure.
Stability. Curvature bounds are stable w.r.t. the natural passage to the limit of the objects which define it.
Interest. Geometrical and analytical consequences on the space can be derived from curvature-dimension condition.

In the next section we recall some basic concepts concerning convergence of metric measure spaces (which are key to discuss the stability issue), while in the following one we give the definition of curvature-dimension condition and analyze its properties.

All the metric measure spaces $(X, d, \boldsymbol{m})$ that we will consider satisfy the following assumption:

Assumption 7.1 *(X, d) is Polish, the measure $\boldsymbol{m}$ is a Borel probability measure and $\boldsymbol{m} \in \mathscr{P}_2(X)$.*

7.1 Convergence of metric measure spaces

We say that two metric measure spaces $(X, d_X, \boldsymbol{m}_X)$ and $(Y, d_Y, \boldsymbol{m}_Y)$ are *isomorphic* provided there exists a bijective isometry $f : \mathrm{supp}(\boldsymbol{m}_X) \to \mathrm{supp}(\boldsymbol{m}_Y)$ such that $f_\#\boldsymbol{m}_X = \boldsymbol{m}_Y$. This is the same as to say that 'we don't care about the behavior of the space (X, d_X) where there is no mass'. This choice will be important in discussing the stability issue.

Definition 7.2 (Coupling between metric measure spaces) *Given two metric measure spaces $(X, d_X, \boldsymbol{m}_X)$, $(Y, d_Y, \boldsymbol{m}_Y)$, we consider the product space $(X \times Y, D_{XY})$, where D_{XY} is the distance defined by*

$$D_{XY}\big((x_1, y_1), (x_2, y_2)\big) := \sqrt{d_X^2(x_1, x_2) + d_Y^2(y_1, y_2)}.$$

We say that a couple $(d, \boldsymbol{\gamma})$ is an admissible coupling between $(X, d_X, \boldsymbol{m}_X)$ and $(Y, d_Y, \boldsymbol{m}_Y)$, we write $(d, \boldsymbol{\gamma}) \in \mathcal{A}dm((d_X, \boldsymbol{m}_X), (d_Y, \boldsymbol{m}_Y))$ if:

- *d is a pseudo distance on $\mathrm{supp}\,\boldsymbol{m}_X \sqcup \mathrm{supp}\,\boldsymbol{m}_Y$ (i.e. it may be zero on two different points) which coincides with d_X (resp. d_Y) when restricted to $\mathrm{supp}\,\boldsymbol{m}_X \times \mathrm{supp}\,\boldsymbol{m}_X$ (resp. $\mathrm{supp}\,\boldsymbol{m}_Y \times \mathrm{supp}\,\boldsymbol{m}_Y$).*
- *a Borel (w.r.t. the Polish structure given by D_{XY}) measure $\boldsymbol{\gamma}$ on $\mathrm{supp}\,\boldsymbol{m}_X \times \mathrm{supp}\,\boldsymbol{m}_Y$ such that $\pi^1_\#\boldsymbol{\gamma} = \boldsymbol{m}_X$ and $\pi^2_\#\boldsymbol{\gamma} = \boldsymbol{m}_Y$.*

It is not hard to see that the set of admissible couplings is always non empty.

The *cost* $C(d, \boldsymbol{\gamma})$ of a coupling is given by

$$C(d, \boldsymbol{\gamma}) := \int_{\mathrm{supp}\,\text{ùscriptsize}\boldsymbol{m}_X \times \mathrm{supp}\,\text{ùscriptsize}\boldsymbol{m}_Y} d^2(x, y)\, d\boldsymbol{\gamma}(x, y).$$

The distance $\mathbb{D}\big((X, d_X, \boldsymbol{m}_X), (Y, d_Y, \boldsymbol{m}_Y)\big)$ is then defined as

$$\mathbb{D}\big((X, d_X, \boldsymbol{m}_X), (Y, d_Y, \boldsymbol{m}_Y)\big) := \inf \sqrt{C(d, \boldsymbol{\gamma})}, \tag{7.4}$$

the infimum being taken among all couplings $(d, \boldsymbol{\gamma})$ of $(X, d_X, \boldsymbol{m}_X)$ and $(Y, d_Y, \boldsymbol{m}_Y)$.

A trivial consequence of the definition is that if $(X, d_X, \boldsymbol{m}_X)$ and $(\tilde{X}, d_{\tilde{X}}, \boldsymbol{m}_{\tilde{X}})$ (resp. $(Y, d_Y, \boldsymbol{m}_Y)$ and $(\tilde{Y}, d_{\tilde{Y}}, \boldsymbol{m}_{\tilde{Y}})$) are isomorphic, then

$$\mathbb{D}\Big((X, d_X, \boldsymbol{m}_X), (Y, d_Y, \boldsymbol{m}_Y)\Big) = \mathbb{D}\Big((\tilde{X}, d_{\tilde{X}}, \boldsymbol{m}_{\tilde{X}}), (\tilde{Y}, d_{\tilde{Y}}, \boldsymbol{m}_{\tilde{Y}})\Big),$$

so that $\mathbb{D}$ is actually defined on isomorphism classes of metric measure spaces.

In the next proposition we collect, without proof, the main properties of $\mathbb{D}$.

Proposition 7.3 (Properties of $\mathbb{D}$) *The* inf *in* (7.4) *is realized, and a coupling realizing it will be called* optimal.

Also, let $\mathbb{X}$ be the set of isomorphism classes of metric measure spaces satisfying Assumption 7.1. Then $\mathbb{D}$ is a distance on $\mathbb{X}$, and in particular $\mathbb{D}$ is 0 only on couples of isomorphic metric measure spaces.

Finally, the space $(\mathbb{X}, \mathbb{D})$ is complete, separable and geodesic.

Proof See Section 3.1 of [74]. □

We will denote by $\mathcal{O}pt((d_X, \boldsymbol{m}_X), (d_Y, \boldsymbol{m}_Y))$ the set of optimal couplings between $(X, d_X, \boldsymbol{m}_X)$ and $(Y, d_Y, \boldsymbol{m}_Y)$, i.e. the set of couplings where the inf in (7.4) is realized.

Given a metric measure space $(X, d, \boldsymbol{m})$ we will denote by $\mathscr{P}_2^a(X) \subset \mathscr{P}(X)$ the set of measures which are absolutely continuous w.r.t. $\boldsymbol{m}$.

To any coupling $(d, \boldsymbol{\gamma})$ of two metric measure spaces $(X, d_X, \boldsymbol{m}_X)$ and $(Y, d_Y, \boldsymbol{m}_Y)$, it is naturally associated a map $\boldsymbol{\gamma}_\# : \mathscr{P}_2^a(X) \to \mathscr{P}_2^a(Y)$ defined as follows:

$$\mu = \rho \boldsymbol{m}_X \qquad \mapsto \qquad \boldsymbol{\gamma}_\# \mu := \eta \boldsymbol{m}_Y, \quad \text{where } \eta \text{ is defined by} \quad \eta(y) := \int \rho(x) d\boldsymbol{\gamma}_y(x), \tag{7.5}$$

where $\{\boldsymbol{\gamma}_y\}$ is the disintegration of $\boldsymbol{\gamma}$ w.r.t. the projection on Y. Similarly, there is a natural map $\boldsymbol{\gamma}_\#^{-1} : \mathscr{P}_2^a(Y) \to \mathscr{P}_2^a(X)$ given by:

$$\nu = \eta \boldsymbol{m}_Y \qquad \mapsto \qquad \boldsymbol{\gamma}_\#^{-1} \nu := \rho \boldsymbol{m}_X, \quad \text{where } \rho \text{ is defined by} \quad \rho(x) := \int \eta(y) d\boldsymbol{\gamma}_x(y),$$

where, obviously, $\{\boldsymbol{\gamma}_x\}$ is the disintegration of $\boldsymbol{\gamma}$ w.r.t. the projection on X.

Notice that $\boldsymbol{\gamma}_\# \boldsymbol{m}_X = \boldsymbol{m}_Y$ and $\boldsymbol{\gamma}_\#^{-1} \boldsymbol{m}_Y = \boldsymbol{m}_X$ and that in general $\boldsymbol{\gamma}_\#^{-1} \boldsymbol{\gamma}_\# \mu \neq \mu$. Also, if $\boldsymbol{\gamma}$ is induced by a map $T : X \to Y$, i.e. if $\boldsymbol{\gamma} = (Id, T)_\# \boldsymbol{m}_X$, then $\boldsymbol{\gamma}_\# \mu = T_\# \mu$ for any $\mu \in \mathscr{P}_2^a(X)$.

Our goal now is to show that if $(X_n, d_n, \boldsymbol{m}_n) \xrightarrow{\mathbb{D}} (X, d, \boldsymbol{m})$ of the *internal energy* kind on $(\mathscr{P}_2^a(X_n), W_2)$ Mosco-converge to the corresponding functional on $(\mathscr{P}_2^a(X), W_2)$. Thus, fix a convex and continuous function $u : [0, +\infty) \to \mathbb{R}$, define

$$u'(\infty) := \lim_{z \to +\infty} \frac{u(z)}{z},$$

and, for every compact metric space (X, d), define the functional $\mathscr{E} : [\mathscr{P}(X)]^2 \to \mathbb{R} \cup \{+\infty\}$ by

$$\mathscr{E}(\mu|\nu) := \int u(\rho) d\nu + u'(\infty) \mu^s(X), \tag{7.6}$$

where $\mu = \rho\nu + \mu^s$ is the decomposition of μ in absolutely continuous $\rho\nu$ and singular part μ^s w.r.t. to ν.

Lemma 7.4 ($\mathscr{E}$ decreases under $\boldsymbol{\gamma}_\#$) *Let $(X, d_X, \boldsymbol{m}_X)$ and $(Y, d_Y, \boldsymbol{m}_Y)$ be two metric measure space and $(d, \boldsymbol{\gamma})$ a coupling between them. Then it holds*

$$\begin{aligned} \mathscr{E}(\boldsymbol{\gamma}_\# \mu | \boldsymbol{m}_Y) &\leq \mathscr{E}(\mu | \boldsymbol{m}_X), \qquad \forall \mu \in \mathscr{P}_2^a(X), \\ \mathscr{E}(\boldsymbol{\gamma}_\#^{-1} \nu | \boldsymbol{m}_X) &\leq \mathscr{E}(\nu | \boldsymbol{m}_Y), \qquad \forall \nu \in \mathscr{P}_2^a(Y). \end{aligned}$$

Proof Clearly it is sufficient to prove the first inequality. Let $\mu = \rho \boldsymbol{m}_X$ and $\boldsymbol{\gamma}_\# \mu = \eta \boldsymbol{m}_Y$, with η given by (7.5). By Jensen's inequality we have

$$\begin{aligned}
\mathscr{E}(\boldsymbol{\gamma}_\# \mu | \boldsymbol{m}_Y) = \int u(\eta(y)) d\boldsymbol{m}_Y(y) &= \int u\left(\int \rho(x) d\boldsymbol{\gamma}_y(x) \right) d\boldsymbol{m}_Y(y) \\
&\leq \int \int u(\rho(x)) d\boldsymbol{\gamma}_y(x) d\boldsymbol{m}_Y(y) = \int u(\rho(x)) d\boldsymbol{\gamma}(x,y) \\
&= \int u(\rho(x)) d\boldsymbol{m}_X(x) = \mathscr{E}(\mu | \boldsymbol{m}_X)
\end{aligned}$$

$\square$

Proposition 7.5 ('Mosco' convergence of internal energy functionals) *Let* $(X_n, d_n, \boldsymbol{m}_n) \overset{\mathbb{D}}{\to} (X, d, \boldsymbol{m})$ *and* $(d_n, \boldsymbol{\gamma}_n) \in \mathcal{O}pt((d_n, \boldsymbol{m}_n), (d, \boldsymbol{m}))$. *Then the following two are true:*

Weak $\boldsymbol{\Gamma} - \underline{\lim}$. *For any sequence* $n \mapsto \mu_n \in \mathscr{P}_2^a(X_n)$ *such that* $n \mapsto (\boldsymbol{\gamma}_n)_\# \mu_n$ *narrowly converges to some* $\mu \in \mathscr{P}(X)$ *it holds*

$$\varliminf_{n\to\infty} \mathscr{E}(\mu_n | \boldsymbol{m}_n) \geq \mathscr{E}(\mu | \boldsymbol{m}).$$

Strong $\boldsymbol{\Gamma} - \overline{\lim}$. *For any* $\mu \in \mathscr{P}_2^a(X)$ *with bounded density there exists a sequence* $n \mapsto \mu_n \in \mathscr{P}_2^a(X_n)$ *such that* $W_2((\boldsymbol{\gamma}_n)_\# \mu_n, \mu) \to 0$ *and*

$$\varlimsup_{n\to\infty} \mathscr{E}(\mu_n | \boldsymbol{m}_n) \leq \mathscr{E}(\mu | \boldsymbol{m}).$$

Note: we put the apexes in *Mosco* because we prove the $\Gamma - \overline{\lim}$ inequality only for measures with bounded densities. This will be enough to prove the stability of Ricci curvature bounds (see Theorem 7.12).

Proof For the first statement we just notice that by Lemma 7.4 we have

$$\mathscr{E}(\mu_n | \boldsymbol{m}_n) \geq \mathscr{E}((\boldsymbol{\gamma}_n)_\# \mu_n | \boldsymbol{m}),$$

and the conclusion follows from the narrow lower semicontinuity of $\mathscr{E}(\cdot | \boldsymbol{m})$.

For the second one we define $\mu_n := (\boldsymbol{\gamma}_n^{-1})_\# \mu$. Then applying Lemma 7.4 twice we get

$$\mathscr{E}(\mu | \boldsymbol{m}) \geq \mathscr{E}(\mu_n | \boldsymbol{m}_n) \geq \mathscr{E}((\boldsymbol{\gamma}_n)_\# \mu_n | \boldsymbol{m}),$$

from which the $\Gamma - \overline{\lim}$ inequality follows. Thus to conclude we need to show that $W_2((\boldsymbol{\gamma}_n)_\# \mu_n, \mu) \to 0$. To check this, we use the Wassertein space built over the (pseudo-)metric space $(X_n \sqcup X, d_n)$: let $\mu = \rho \boldsymbol{m}_X$ and for any $n \in \mathbb{N}$ define the plan $\tilde{\boldsymbol{\gamma}}_n \in \mathscr{P}(X_n \times X)$ by $d\tilde{\boldsymbol{\gamma}}_n(y,x) := \rho(x) d\boldsymbol{\gamma}_n(y,x)$ and notice that $\tilde{\boldsymbol{\gamma}}_n \in \mathcal{A}dm(\mu_n, \mu)$. Thus

$$W_2(\mu_n, \mu) \leq \sqrt{\int d_n^2(x,y) d\tilde{\boldsymbol{\gamma}}_n(y,x)} \leq \sqrt{\int d_n^2(x,y) \rho(x) d\boldsymbol{\gamma}_n(y,x)} \leq \sqrt{M} \sqrt{C(d_n, \boldsymbol{\gamma}_n)},$$

where M is the essential supremum of ρ. By definition, it is immediate to check that the density η_n of μ_n is also bounded above by M. Introduce the plan $\overline{\boldsymbol{\gamma}}_n$ by $d\overline{\boldsymbol{\gamma}}_n(y,x) := \eta_n(y) d\boldsymbol{\gamma}_n(y,x)$ and notice that $\overline{\boldsymbol{\gamma}}_n \in \mathcal{A}dm(\mu_n, (\boldsymbol{\gamma}_n)_\# \mu_n)$, so that, as before, we have

$$W_2(\mu_n, (\boldsymbol{\gamma}_n)_\# \mu_n) \leq \sqrt{\int d_n^2(x,y) d\overline{\boldsymbol{\gamma}}_n(y,x)} \leq \sqrt{\int d_n^2(x,y) \eta_n(y) d\boldsymbol{\gamma}_n(y,x)} \leq \sqrt{M} \sqrt{C(d_n, \boldsymbol{\gamma}_n)}.$$

In conclusion we have

$$W_2(\mu, (\boldsymbol{\gamma}_n)_\# \mu_n) \leq W_2(\mu_n, (\boldsymbol{\gamma}_n)_\# \mu_n) + W_2(\mu_n, \mu) \leq 2\sqrt{M} \sqrt{C(d_n, \boldsymbol{\gamma}_n)},$$

which gives the thesis. $\square$

7.2 Weak Ricci curvature bounds: definition and properties

Define the functions u_N, $N > 1$, and u_∞ on $[0, +\infty)$ as

$$u_N(z) := N(z - z^{1-\frac{1}{N}}),$$

and

$$u_\infty(z) := z \log(z).$$

Then given a metric measure space $(X, d, \boldsymbol{m})$ we define the functionals $\mathscr{E}_N, \mathscr{E}_\infty : \mathscr{P}(X) \to \mathbb{R} \cup \{+\infty\}$ by

$$\mathscr{E}_N(\mu) := \mathscr{E}(\mu|\boldsymbol{m}),$$

where $\mathscr{E}(\cdot|\cdot)$ is given by formula (7.6) with $u := u_N$; similarly for $\mathscr{E}_\infty$.

The definitions of weak Ricci curvature bounds are the following:

Definition 7.6 (Curvature $\geq K$ and no bound on dimension - $CD(K, \infty)$) *We say that a metric measure space $(X, d, \boldsymbol{m})$ has Ricci curvature bounded from below by $K \in \mathbb{R}$ provided the functional*

$$\mathscr{E}_\infty : \mathscr{P}(X) \to \mathbb{R} \cup \{+\infty\},$$

is K-geodesically convex on $(\mathscr{P}_2^a(X), W_2)$. In this case we say that $(X, d, \boldsymbol{m})$ satisfies the curvature dimension condition $CD(K, \infty)$ or that $(X, d, \boldsymbol{m})$ is a $CD(K, \infty)$ space.

Definition 7.7 (Curvature ≥ 0 and dimension $\leq N$ - $CD(0, N)$) *We say that a metric measure space $(X, d, \boldsymbol{m})$ has nonnegative Ricci curvature and dimension bounded from above by N provided the functionals*

$$\mathscr{E}_{N'} : \mathscr{P}(X) \to \mathbb{R} \cup \{+\infty\},$$

are geodesically convex on $(\mathscr{P}_2^a(X), W_2)$ for every $N' \geq N$. In this case we say that $(X, d, \boldsymbol{m})$ satisfies the curvature dimension condition $CD(0, N)$, or that $(X, d, \boldsymbol{m})$ is a $CD(0, N)$ space.

Note that $N > 1$ is not necessarily an integer.

Remark 7.8 Notice that geodesic convexity is required on $\mathscr{P}_2(\mathrm{supp}(\boldsymbol{m}_X))$ and not on $\mathscr{P}_2(X)$. This makes no difference for what concerns $CD(K, \infty)$ spaces, as $\mathscr{E}_\infty$ is $+\infty$ on measures having a singular part w.r.t. $\boldsymbol{m}$, but is important for the case of $CD(0, N)$ spaces, as the functional $\mathscr{E}_N$ has only real values, and requiring geodesic convexity on the whole $\mathscr{P}_2(X)$ would lead to a notion not invariant under isomorphism of metric measure spaces.

Also, for the $CD(0, N)$ condition one requires the geodesic convexity of all $\mathscr{E}_{N'}$ to ensure the following compatibility condition: if X is a $CD(0, N)$ space, then it is also a $CD(0, N')$ space for any $N' > N$. Using Proposition 2.16 it is not hard to see that such compatibility condition is automatically satisfied on non branching spaces. ■

Remark 7.9 (How to adapt the definitions to general bounds on curvature the dimension) It is pretty natural to guess that the notion of bound from below on the Ricci curvature by $K \in \mathbb{R}$ and bound from above on the dimension by N can be given by requiring the functional $\mathscr{E}_N$ to be K-geodesically convex on $(\mathscr{P}(X), W_2)$. However, this is *wrong*, because such condition is not compatible with the Riemannian case. The hearth of the definition of $CD(K, N)$ spaces still concerns the properties of $\mathscr{E}_N$, but a different and more complicated notion of "convexity" is involved. ■

Let us now check that the definitions given have the qualitative properties that we discussed in the introduction of this chapter.
Intrinsicness. This property is clear from the definition.
Compatibility. To give the answer we need to do some computations on Riemannian manifolds:

Lemma 7.10 (Second derivative of the internal energy) *Let M be a compact and smooth Riemannian manifold, $\boldsymbol{m}$ its normalized volume measure, $u : [0,+\infty)$ be convex, continuous and C^2 on $(0,+\infty)$ with $u(0)=0$ and define the "pressure" $p:[0,+\infty)\to\mathbb{R}$ by*

$$p(z) := zu'(z) - u(z), \qquad \forall z > 0,$$

and $p(0) := 0$. Also, let $\mu = \rho\boldsymbol{m} \in \mathscr{P}_2^a(M)$ with $\rho \in C^\infty(M)$, pick $\varphi \in C_c^\infty(M)$, and define $T_t : M \to M$ by $T_t(x) := \exp_x(t\nabla\varphi(x))$. Then it holds:

$$\frac{d^2}{dt^2}\Big|_{t=0}\mathscr{E}((T_t)_\#\mu) = \int p'(\rho)\,\rho\,(\Delta\varphi)^2 - p(\rho)\Big((\Delta\varphi)^2 - \big|\nabla^2\varphi\big|^2 - \mathrm{Ric}\big(\nabla\varphi,\nabla\varphi\big)\Big)\,d\boldsymbol{m},$$

where by $\big|\nabla^2\varphi(x)\big|^2$ we mean the trace of the linear map $(\nabla^2\varphi(x))^2 : T_xM \to T_xM$ (in coordinates, this reads as $\sum_{ij}(\partial_{ij}\varphi(x))^2$).

Proof
(Computation of the second derivative). Let $D_t(x) := \det(\nabla T_t(x))$, $\mu_t := (T_t)_\#\mu = \rho_t\mathrm{Vol}$. By compactness, for t sufficiently small T_t is invertible with smooth inverse, so that $D_t, \rho_t \in C^\infty(M)$. For small t, the change of variable formula gives

$$\rho_t(T_t(x)) = \frac{\rho(x)}{\det(\nabla T_t(x))} = \frac{\rho(x)}{D_t(x)}.$$

Thus we have (all the integrals being w.r.t. $\boldsymbol{m}$):

$$\frac{d}{dt}\int u(\rho_t) = \frac{d}{dt}\int u\left(\frac{\rho}{D_t}\right)D_t = \int -u'\left(\frac{\rho}{D_t}\right)\frac{\rho D_t'}{D_t^2}D_t + u\left(\frac{\rho}{D_t}\right)D_t' = -\int p\left(\frac{\rho}{D_t}\right)D_t',$$

and

$$\frac{d^2}{dt^2}\Big|_{t=0}\int u(\rho_t) = -\frac{d}{dt}\Big|_{t=0}\int p\left(\frac{\rho}{D_t}\right)D_t' = \int p'(\rho)\rho(D_0')^2 - p(\rho)D_0'',$$

having used the fact that $D_0 \equiv 1$.
(Evaluation of D_0' and D_0''). We want to prove that

$$\begin{aligned} D_0'(x) &= \Delta\varphi(x), \\ D_0''(x) &= (\Delta\varphi(x))^2 - \big|\nabla^2\varphi(x)\big|^2 - \mathrm{Ric}\big(\nabla\varphi(x),\nabla\varphi(x)\big). \end{aligned} \tag{7.7}$$

For $t \geq 0$ and $x \in M$, let $J_t(x)$ be the operator from T_xM to $T_{\exp_x(t\nabla\varphi(x))}M$ given by:

$$J_t(x)(v) := \begin{cases} \text{the value at } s=t \text{ of the Jacobi field } j_s \text{ along the geodesic} \\ s\mapsto \exp_x(s\nabla\varphi(x)), \text{having the initial conditions } j_0 := v,\ j_0' := \nabla^2\varphi\cdot v, \end{cases}$$

(where here and in the following the apex $'$ on a vector/tensor field stands for covariant differentiation), so that in particular we have

$$\begin{aligned} J_0 &= Id, \\ J_0' &= \nabla^2\varphi. \end{aligned} \tag{7.8}$$

The fact that Jacobi fields are the differential of the exponential map reads, in our case, as:

$$\nabla T_t(x)\cdot v = J_t(x)\cdot v,$$

therefore we have

$$D_t = \det(J_t). \tag{7.9}$$

Also, Jacobi fields satisfy the Jacobi equation, which we write as

$$J_t'' + A_t J_t = 0, \tag{7.10}$$

where $A_t(x) : T_{\exp_x(t\nabla\varphi(x))}M \to T_{\exp_x(t\nabla\varphi(x))}M$ is the map given by

$$A_t(x)\cdot v := R(\dot\gamma_t, v)\dot\gamma_t,$$

where $\gamma_t := \exp_x(t\nabla\varphi(x))$. Recalling the rule $(\det B_t)' = \det(B_t)\mathrm{tr}(B_t'B_t^{-1})$, valid for a smooth curve of linear operators, we obtain from (7.9) the validity of

$$D_t' = D_t\mathrm{tr}(J_t'J_t^{-1}). \tag{7.11}$$

Evaluating this identity at $t=0$ and using (7.8) we get the first of (7.7). Recalling the rule $(B_t^{-1})' = -B_t^{-1}B_t'B_t^{-1}$, valid for a smooth curve of linear operators, and differentiating in time equation (7.11) we obtain

$$D_t'' = D_t\big(\mathrm{tr}(J_t'J_t^{-1})\big)^2 + D_t\mathrm{tr}(J_t''J_t^{-1} - J_t'J_t^{-1}J_t'J_t^{-1}) = D_t\Big(\big(\mathrm{tr}(J_t'J_t^{-1})\big)^2 - \mathrm{tr}\big(A_t + J_t'J_t^{-1}J_t'J_t^{-1}\big)\Big),$$

having used the Jacobi equation (7.10). Evaluate this expression at $t=0$, use (7.8) and observe that

$$\mathrm{tr}(A_0) = \mathrm{tr}\Big\{v \mapsto R(\nabla\varphi, v)\nabla\varphi\Big\} = \mathrm{Ric}(\nabla\varphi,\nabla\varphi),$$

to get the second of (7.7). □

Theorem 7.11 (Compatibility of weak Ricci curvature bounds) *Let M be a compact Riemannian manifold, d its Riemannian distance and $\boldsymbol{m}$ its normalized volume measure. Then:*

- *i) the functional $\mathscr{E}_\infty$ is K-geodesically convex on $(\mathscr{P}_2(M), W_2)$ if and only if M has Ricci curvature uniformly bounded from below by K.*
- *ii) the functional $\mathscr{E}_N$ is geodesically convex on $(\mathscr{P}_2(M), W_2)$ if and only if M has non negative Ricci curvature and $\dim(M) \le N$.*

Sketch of the Proof We will give only a formal proof, neglecting all the issues which arise due to the potential non regularity of the objects involved.

We start with (i). Assume that $\mathrm{Ric}(v,v) \ge K|v|^2$ for any v. Pick a geodesic $(\rho_t\boldsymbol{m}) \subset \mathscr{P}_2(M)$ and assume that $\rho_t \in C^\infty$ for any $t\in[0,1]$. By Theorem 1.33 we know that there exists a function $\varphi : M \to \mathbb{R}$ differentiable $\rho_0\boldsymbol{m}$-a.e. such that $\exp(\nabla\varphi)$ is the optimal transport map from $\rho_0\boldsymbol{m}$ to $\rho_1\boldsymbol{m}$ and

$$\rho_t\boldsymbol{m} = \big(\exp(t\nabla\varphi)\big)_\#\rho_0\boldsymbol{m}.$$

Assume that φ is C^∞. Then by Lemma 7.10 with $u := u_\infty$ we know that

$$\frac{d^2}{dt^2}\mathscr{E}_\infty(\rho_t\boldsymbol{m}) = \int \Big(\big|\nabla^2\varphi\big|^2 + \mathrm{Ric}(\nabla\varphi,\nabla\varphi)\Big)\rho_0\,d\boldsymbol{m} \ge K\int |\nabla\varphi|^2\rho_0\,d\boldsymbol{m}.$$

Since $\int |\nabla\varphi|^2\rho_0 d\boldsymbol{m} = W_2^2(\rho_0,\rho_1)$, the claim is proved.

The converse implication follows by an explicit construction: if $\mathrm{Ric}(v,v) < K|v|^2$ for some $x \in M$ and $v \in T_xM$, then for $\varepsilon \ll \delta \ll 1$ define $\mu_0 := c_0\boldsymbol{m}|_{B_\varepsilon(x)}$ (c_0 being the normalizing constant) and $\mu_t := (T_t)_\#\mu_0$ where $T_t(y) := \exp_y(t\delta\nabla\varphi(y))$ and $\varphi \in C^\infty$ is such that $\nabla\varphi(x) = v$ and $\nabla^2\varphi(x) = 0$. Using Lemma 7.10 again and the hypothesis $\mathrm{Ric}(v,v) < K|v|^2$ it is not hard to prove that $\mathscr{E}_\infty$ is not λ-geodesically convex along (μ_t). We omit the details.

Now we turn to (ii). Let $(\rho_t\boldsymbol{m})$ and φ as in the first part of the argument above. Assume that M has non negative Ricci curvature and that $\dim(M) \le N$. Observe that for $u := u_N$ Lemma 7.10 gives

$$\frac{d^2}{dt^2}|_{t=0}\mathscr{E}_N(\rho_t) = \int \left(1-\frac{1}{N}\right)\rho^{1-\frac{1}{N}}(\Delta\varphi)^2 - \rho^{1-\frac{1}{N}}\left((\Delta\varphi)^2 - \left|\nabla^2\varphi\right|^2 - \frac{1}{2}\mathrm{Ric}(\nabla\varphi,\nabla\varphi)\right)d\boldsymbol{m}.$$

Using the hypothesis on M and the fact that $(\Delta\varphi)^2 \le N\left|\nabla^2\varphi\right|^2$ we get $\frac{d^2}{dt^2}|_{t=0}\mathscr{E}_N(\rho_t) \ge 0$, i.e. the geodesic convexity of $\mathscr{E}_N$. For the converse implication it is possible to argue as above, we omit the details also in this case. $\square$

Now we pass to the **stability**:

Theorem 7.12 (Stability of weak Ricci curvature bound) *Assume that* $(X_n,d_n,\boldsymbol{m}_n) \overset{\mathbb{D}}{\to} (X,d,\boldsymbol{m})$ *and that for every* $n \in \mathbb{N}$ *the space* $(X_n,d_n,\boldsymbol{m}_n)$ *is* $CD(K,\infty)$ *(resp.* $CD(0,N)$*). Then* $(X,d,\boldsymbol{m})$ *is a* $CD(K,\infty)$ *(resp.* $CD(0,N)$*) space as well.*

Sketch of the Proof Pick $\mu_0,\mu_1 \in \mathscr{P}_2^a(X)$ and assume they are both absolutely continuous with bounded densities, say $\mu_i = \rho_i\boldsymbol{m}$, $i=0,1$. Choose $(\tilde{d}_n,\boldsymbol{\gamma}_n) \in \mathcal{O}pt((d_n,\boldsymbol{m}_n),(d,\boldsymbol{m}))$. Define $\mu_i^n := (\boldsymbol{\gamma}_n^{-1})_\#\mu_i \in \mathscr{P}_2^a(X_n)$, $i=0,1$. Then by assumption there is a geodesic $(\mu_t^n) \subset \mathscr{P}_2^a(X_n)$ such that

$$\mathscr{E}_\infty(\mu_t^n) \le (1-t)\mathscr{E}_\infty(\mu_0^n) + t\mathscr{E}_\infty(\mu_1^n) - \frac{K}{2}t(1-t)W_2^2(\mu_0^n,\mu_1^n). \tag{7.12}$$

Now let $\sigma_t^n := (\boldsymbol{\gamma}_n)_\#\mu_t^n \in \mathscr{P}_2^a(X)$, $t \in [0,1]$. From Proposition 7.5 and its proof we know that $W_2(\mu_i,\sigma_i^n) \to 0$ as $n \to \infty$, $i=0,1$. Also, from (7.12) ad Lemma 7.4, we know that $\mathscr{E}_\infty(\sigma_t^n)$ is uniformly bounded in n,t. Thus for every fixed t the sequence $n \mapsto \sigma_t^n$ is tight, and we can extract a subsequence, not relabeled, such that σ_t^n narrowly converges to some $\sigma_t \in \mathscr{P}_2(\mathrm{supp}(\boldsymbol{m}))$ for every rational t. By an equicontinuity argument it is not hard to see that then σ_t^n narrowly converges to some σ_t for any $t \in [0,1]$ (we omit the details). We claim that (σ_t) is a geodesic, and that the K-convexity inequality is satisfied along it. To check that it is a geodesic just notice that for any partition $\{t_i\}$ of $[0,1]$ we have

$$\begin{aligned}W_2(\mu_0,\mu_1) &= \lim_{n\to\infty} W_2(\sigma_0^n,\sigma_1^n) = \lim_{n\to\infty}\sum_i W_2(\sigma_{t_i}^n,\sigma_{t_{i+1}}^n)\\ &\ge \sum_i \varliminf_{n\to\infty} W_2(\sigma_{t_i}^n,\sigma_{t_{i+1}}^n) \ge \sum_i W_2(\sigma_{t_i},\sigma_{t_{i+1}}).\end{aligned}$$

Passing to the limit in (7.12), recalling Proposition 7.5 to get that $\mathscr{E}_\infty(\mu_i^n) \to \mathscr{E}_\infty(\mu_i)$, $i=0,1$, and that $\varliminf_{n\to\infty}\mathscr{E}_\infty(\mu_t^n) \ge \varliminf_{n\to\infty}\mathscr{E}_\infty(\sigma_t^n) \ge \mathscr{E}_\infty(\sigma_t)$ we conclude.

To deal with general μ_0,μ_1, we start recalling that the sublevels of $\mathscr{E}_\infty$ are tight, indeed using first the bound $z\log(z) \ge -\frac{1}{e}$ and then Jensen's inequality we get

$$\frac{1}{e} + C \ge \frac{\boldsymbol{m}(X\setminus E)}{e} + \mathscr{E}_\infty(\mu) \ge \int_E \rho\log(\rho)d\boldsymbol{m} \ge \mu(E)\log\left(\frac{\mu(E)}{\boldsymbol{m}(E)}\right),$$

for any $\mu = \rho \boldsymbol{m}$ such that $\mathscr{E}_\infty(\mu) \leq C$ and any Borel $E \subset X$. This bound gives that if $\boldsymbol{m}(E_n) \to 0$ then $\mu(E_n) \to 0$ uniformly on the set of μ's such that $\mathscr{E}_\infty(\mu) \leq C$. This fact together with the tightness of $\boldsymbol{m}$ gives the claimed tightness of the sublevels of $\mathscr{E}_\infty$.

Now the conclusion follows by a simple truncation argument using the narrow compactness of the sublevels of $\mathscr{E}_\infty$ and the lower semicontinuity of $\mathscr{E}_\infty$ w.r.t. narrow convergence.

For the stability of the $CD(0, N)$ condition, the argument is the following: we first deal with the case of μ_0, μ_1 with bounded densities with exactly the same ideas used for $\mathscr{E}_\infty$. Then to pass to the general case we use the fact that if $(X, d, \boldsymbol{m})$ is a $CD(0, N)$ space, then $(\mathrm{supp}(\boldsymbol{m}), d, \boldsymbol{m})$ is a doubling space (Proposition 7.15 below - notice that $\mathscr{E}_{N'} \leq N'$ and thus it is not true that sublevels of $\mathscr{E}_{N'}$ are tight) and therefore boundedly compact. Then the inequality

$$R^2 \mu(\mathrm{supp}(\boldsymbol{m}) \setminus B_R(x_0)) \leq \int d^2(\cdot, x_0) d\mu,$$

shows that the set of μ's in $\mathscr{P}_2^a(X)$ with bounded second moment is tight. Hence the conclusion follows, as before, using this narrow compactness together with the lower semicontinuity of $\mathscr{E}_{N'}$ w.r.t. narrow convergence. $\square$

It remains to discuss the **interest**: from now on we discuss some of the geometric and analytic properties of spaces having a weak Ricci curvature bound.

Proposition 7.13 (Restriction and rescaling) *Let $(X, d, \boldsymbol{m})$ be a $CD(K, \infty)$ space (resp. $CD(0, N)$ space). Then:*

i) **Restriction**. *If $Y \subset X$ is a closed totally convex subset (i.e. every geodesic with endpoints in Y lies entirely inside Y) such that $\boldsymbol{m}(Y) > 0$, then the space $(Y, d, \boldsymbol{m}(Y)^{-1}\boldsymbol{m}_{|_Y})$ is a $CD(K, \infty)$ space (resp. $CD(0, N)$ space),*

ii) **Rescaling**. *for every $\alpha > 0$ the space $(X, \alpha d, \boldsymbol{m})$ is a $CD(\alpha^{-2}K, \infty)$ space (resp. $CD(0, N)$ space).*

Proof
(**i**). Pick $\mu_0, \mu_1 \in \mathscr{P}(Y) \subset \mathscr{P}(X)$ and a constant speed geodesic $(\mu_t) \subset \mathscr{P}(X)$ connecting them such that

$$\mathscr{E}_\infty(\mu_t) \leq (1-t)\mathscr{E}_\infty(\mu_0) + t\mathscr{E}_\infty(\mu_1) - \frac{K}{2} t(1-t) W_2^2(\mu_0, \mu_1),$$

(resp. satisfying the convexity inequality for the functional $\mathscr{E}_{N'}$, $N' \geq N$).

We claim that $\mathrm{supp}(\mu_t) \subset Y$ for any $t \in [0, 1]$. Recall Theorem 2.10 and pick a measure $\boldsymbol{\mu} \in \mathscr{P}(\mathrm{Geod}(X))$ such that

$$\mu_t = (\mathrm{e}_t)_\# \boldsymbol{\mu},$$

where e_t is the evaluation map defined by equation (2.6). Since $\mathrm{supp}(\mu_0), \mathrm{supp}(\mu_1) \subset Y$ we know that for any geodesic $\gamma \in \mathrm{supp}(\boldsymbol{\mu})$ it holds $\gamma_0, \gamma_1 \in Y$. Since Y is totally convex, this implies that $\gamma_t \in Y$ for any t and any $\gamma \in \mathrm{supp}(\boldsymbol{\mu})$, i.e. $\mu_t = (\mathrm{e}_t)_\# \boldsymbol{\mu} \in \mathscr{P}(Y)$. Therefore (μ_t) is a geodesic connecting μ_0 to μ_1 in (Y, d). Conclude noticing that for any $\mu \in \mathscr{P}_2(Y)$ it holds

$$\int \frac{d\mu}{d\boldsymbol{m}_Y} \log\left(\frac{d\mu}{d\boldsymbol{m}_Y}\right) d\boldsymbol{m}_Y = \log(\boldsymbol{m}(Y)) + \int \frac{d\mu}{d\boldsymbol{m}} \log\left(\frac{d\mu}{d\boldsymbol{m}}\right) d\boldsymbol{m},$$

$$\int \left(\frac{d\mu}{d\boldsymbol{m}_Y}\right)^{1-\frac{1}{N'}} d\boldsymbol{m}_Y = \boldsymbol{m}(Y)^{-\frac{1}{N'}} \int \left(\frac{d\mu}{d\boldsymbol{m}}\right)^{1-\frac{1}{N'}} d\boldsymbol{m},$$

where we wrote $\boldsymbol{m}_Y$ for $\boldsymbol{m}(Y)^{-1}\boldsymbol{m}_{|_Y}$.
(**ii**). Fix $\alpha > 0$ and let $\tilde{d} := \alpha d$ and $\tilde{W}_2$ be the Wasserstein distance on $\mathscr{P}(X)$ induced by the

distance $\tilde{d}$. It is clear that a plan $\boldsymbol{\gamma} \in \mathcal{Adm}(\mu, \nu)$ is optimal for the distance W_2 if and only if it is optimal for $\tilde{W}_2$, thus $\tilde{W}_2 = \alpha W_2$. Now pick $\mu_0,\ \mu_1 \in \mathscr{P}(X)$ and let $(\mu_t) \subset \mathscr{P}(X)$ be a constant speed geodesic connecting them such that

$$\mathscr{E}_\infty(\mu_t) \leq (1-t)\mathscr{E}(\mu_0) + t\mathscr{E}(\mu_1) - \frac{K}{2}t(1-t)W_2^2(\mu_0, \mu_1),$$

then it holds

$$\mathscr{E}_\infty(\mu_t) \leq (1-t)\mathscr{E}(\mu_0) + t\mathscr{E}(\mu_1) - \frac{K}{2\alpha^2}t(1-t)\tilde{W}_2^2(\mu_0, \mu_1),$$

and the proof is complete. A similar argument applies for the case $CD(0, N)$. □

For $A_0, A_1 \subset X$, we define $[A_0, A_1]_t \subset X$ as:

$$[A_0, A_1]_t := \Big\{ \gamma(t)\ :\ \gamma \text{ is a constant speed geodesic such that } \gamma(0) \in A_0,\ \gamma(1) \in A_1 \Big\}.$$

Observe that if A_0, A_1 are open (resp. compact) $[A_0, A_1]_t$ is open (resp. compact), hence Borel.

Proposition 7.14 (Brunn-Minkowski) *Let $(X, d, \boldsymbol{m})$ be a metric measure space and $A_0,\ A_1 \subset \operatorname{supp}(\boldsymbol{m})$ compact subsets. Then:*

$i)$ if $(X, d, \boldsymbol{m})$ is a $CD(K, \infty)$ space it holds:

$$\log(\boldsymbol{m}([A_0, A_1]_t)) \geq (1-t)\log(\boldsymbol{m}(A_0)) + t\log(\boldsymbol{m}(A_1)) + \frac{K}{2}t(1-t)D_K^2(A_0, A_1), \quad (7.13)$$

where $D_K(A_0, A_1)$ is defined as $\sup_{\substack{x_0 \in A_0\\ x_1 \in A_1}} d(x_0, x_1)$ if $K < 0$ and as $\inf_{\substack{x_0 \in A_0\\ x_1 \in A_1}} d^2(x_0, x_1)$ if $K > 0$.

$ii)$ If $(X, d, \boldsymbol{m})$ is a $CD(0, N)$ space it holds:

$$\boldsymbol{m}\big([A_0, A_1]_t\big)^{1/N} \geq (1-t)\boldsymbol{m}(A_0)^{1/N} + t\boldsymbol{m}(A_1)^{1/N}. \quad (7.14)$$

Proof We start with (i). Suppose that A_0, A_1 are open satisfying $\boldsymbol{m}(A_0), \boldsymbol{m}(A_1) > 0$. Define the measures $\mu_i := \boldsymbol{m}(A_i)^{-1}\boldsymbol{m}_{|_{A_i}}$ for $i = 0, 1$ and find a constant speed geodesic $(\mu_t) \subset \mathscr{P}(X)$ such that

$$\mathscr{E}_\infty(\mu_t) \leq (1-t)\mathscr{E}_\infty(\mu_0) + t\mathscr{E}_\infty(\mu_1) - \frac{K}{2}t(1-t)W_2^2(\mu_0, \mu_1).$$

Arguing as in the proof of the previous proposition, it is immediate to see that μ_t is concentrated on $[A_0, A_1]_t$ for any $t \in [0, 1]$.

In particular $\boldsymbol{m}([A_0, A_1]_t) > 0$, otherwise $\mathscr{E}_\infty(\mu_t)$ would be $+\infty$ and the convexity inequality would fail. Now let $\nu_t := \boldsymbol{m}([A_0, A_1]_t)^{-1}\boldsymbol{m}_{|_{[A_0, A_1]_t}}$: an application of Jensen inequality shows that $\mathscr{E}_\infty(\mu_t) \geq \mathscr{E}_\infty(\nu_t)$, thus we have

$$\mathscr{E}_\infty(\nu_t) \leq (1-t)\mathscr{E}_\infty(\mu_0) + t\mathscr{E}_\infty(\mu_1) - \frac{K}{2}t(1-t)W_2^2(\mu_0, \mu_1).$$

Notice that for a general μ of the form $\boldsymbol{m}(A)^{-1}\boldsymbol{m}_{|_A}$ it holds

$$\mathscr{E}_\infty(\mu) = \log\big(\boldsymbol{m}(A)^{-1}\big) = -\log\big(\boldsymbol{m}(A)\big),$$

and conclude using the trivial inequality

$$\inf_{\substack{x_0 \in A_0\\ x_1 \in A_1}} d^2(x_0, x_1) \leq W_2^2(\mu_0, \mu_1) \leq \sup_{\substack{x_0 \in A_0\\ x_1 \in A_1}} d^2(x_0, x_1).$$

The case of A_0, A_1 compact now follows by a simple approximation argument by considering the ε-neighborhood $A_i^\varepsilon := \{x : d(x, A_i) < \varepsilon\}$, $i = 0, 1$, noticing that $[A_0, A_1]_t = \cap_{\varepsilon>0}[A_0^\varepsilon, A_1^\varepsilon]_t$, for any $t \in [0,1]$ and that $\boldsymbol{m}(A_i^\varepsilon) > 0$ because $A_i \subset \operatorname{supp}(\boldsymbol{m})$, $i = 0, 1$.

Part (ii) follows along the same lines taking into account that for a general μ of the form $\boldsymbol{m}(A)^{-1}\boldsymbol{m}_{|A}$ it holds

$$\mathscr{E}_N(\mu) = N(1 - \boldsymbol{m}(A)^{1/N}),$$

and that, as before, if $\boldsymbol{m}(A_0), \boldsymbol{m}(A_1) > 0$ it cannot be $\boldsymbol{m}([A_0, A_1]_t) = 0$ or we would violate the convexity inequality. □

A consequence of Brunn-Minkowski is the Bishop-Gromov inequality.

Proposition 7.15 (Bishop-Gromov) *Let $(X, d, \boldsymbol{m})$ be a $CD(0, N)$ space. Then it holds*

$$\frac{\boldsymbol{m}(B_r(x))}{\boldsymbol{m}(B_R(x))} \geq \left(\frac{r}{R}\right)^N, \qquad \forall x \in \operatorname{supp}(\boldsymbol{m}). \tag{7.15}$$

In particular, $(\operatorname{supp}(\boldsymbol{m}), d, \boldsymbol{m})$ *is a doubling space.*

Proof Pick $x \in \operatorname{supp}(\boldsymbol{m})$ and assume that $\boldsymbol{m}(\{x\}) = 0$. Let $v(r) := \boldsymbol{m}(\overline{B_r(x)})$. Fix $R > 0$ and apply the Brunn-Minkowski inequality to $A_0 = \{x\}$, $A_1 = B_R(x)$ observing that $[A_0, A_1]_t \subset \overline{B_{tR}(x)}$ to get

$$v^{1/N}(tR) \geq \boldsymbol{m}\big([A_0, A_1]_t\big)^{1/N} \geq t v^{1/N}(R), \qquad \forall 0 \leq t \leq 1.$$

Now let $r := tR$ and use the arbitrariness of R, t to get the conclusion.

It remains to deal with the case $\boldsymbol{m}(\{x\}) \neq 0$. We can also assume $\operatorname{supp}(\boldsymbol{m}) \neq \{x\}$, otherwise the thesis would be trivial: under this assumption we will prove that $\boldsymbol{m}(\{x\}) = 0$ for any $x \in X$.

A simple consequence of the geodesic convexity of $\mathscr{E}_N$ tested with delta measures is that $\operatorname{supp}(\boldsymbol{m})$ is a geodesically convex set, therefore it is uncountable. Then there must exist some $x' \in \operatorname{supp}(\boldsymbol{m})$ such that $\boldsymbol{m}(\{x'\}) = 0$. Apply the previous argument with x' in place of x to get that

$$\frac{v(r)}{v(R)} \geq \left(\frac{r}{R}\right)^N, \qquad \forall 0 \leq r < R, \tag{7.16}$$

where now $v(r)$ is the volume of the closed ball of radius r around x'. By definition, v is right continuous; letting $r \uparrow R$ we obtain from (7.16) that v is also left continuous. Thus it is continuous, and in particular the volume of the spheres $\{y : d(y, x') = r\}$ is 0 for any $r \geq 0$. In particular $\boldsymbol{m}(\{y\}) = 0$ for any $y \in X$ and the proof is concluded. □

An interesting geometric consequence of the Brunn-Minkowski inequality in conjunction with the non branching hypothesis is the fact that the 'cut-locus' is negligible.

Proposition 7.16 (Negligible cut-locus) *Assume that $(X, d, \boldsymbol{m})$ is a $CD(0, N)$ space and that it is non branching. Then for every $x \in \operatorname{supp}(\boldsymbol{m})$ the set of y's such that there is more than one geodesic from x to y is $\boldsymbol{m}$-negligible. In particular, for $\boldsymbol{m} \times \boldsymbol{m}$-a.e. (x, y) there exists only one geodesic $\gamma^{x,y}$ from x to y and the map $X^2 \ni (x, y) \mapsto \gamma^{x,y} \in \operatorname{Geod}(X)$ is measurable.*

Proof Fix $x \in \operatorname{supp}(\boldsymbol{m})$, $R > 0$ and consider the sets $A_t := [\{x\}, B_R(x)]_t$. Fix $t < 1$ and $y \in A_t$. We claim that there is only one geodesic connecting it to x. By definition, we know that there is some $z \in B_R(x)$ and a geodesic γ from z to x such that $\gamma_t = y$. Now argue by contradiction and assume that there are 2 geodesics γ^1, γ^2 from y to x. Then starting from z, following γ for time $1 - t$, and then following each of γ^1, γ^2 for the rest of the time we find 2 different geodesics from z to x which agree on the non trivial interval $[0, 1-t]$. This contradicts the non-branching hypothesis.

Clearly $A_t \subset A_s \subset B_R(x)$ for $t \leq s$, thus $t \mapsto \boldsymbol{m}(A_t)$ is non decreasing. By (7.14) and the fact that $\boldsymbol{m}(\{x\}) = 0$ (proved in Proposition 7.15) we know that $\lim_{t\to 1} \boldsymbol{m}(A_t) = \boldsymbol{m}(B_R(x))$ which means that $\boldsymbol{m}$-a.e. point in $B_R(x)$ is connected to x by a unique geodesic. Since R and x are arbitrary, uniqueness is proved.

The measurability of the map $(x, y) \mapsto \gamma^{x,y}$ is then a consequence of uniqueness, of Lemma 2.11 and classical measurable selection results, which ensure the existence of a measurable selection of geodesics: in our case there is $\boldsymbol{m} \times \boldsymbol{m}$-almost surely no choice, so the unique geodesic selection is measurable. □

Corollary 7.17 (Compactness) *Let N, $D < \infty$. Then the family $\mathcal{X}(N, D)$ of (isomorphism classes of) metric measure spaces $(X, d, \boldsymbol{m})$ satisfying the condition $CD(0, N)$, with diameter bounded above by D is compact w.r.t. the topology induced by $\mathbb{D}$.*

Sketch of the Proof Using the Bishop-Gromov inequality with $R = D$ we get that

$$\boldsymbol{m}(\overline{B_\varepsilon(x)}) \geq \left(\frac{\varepsilon}{D}\right)^N, \qquad \forall (X, d, \boldsymbol{m}) \in \mathcal{X}(N, D),\ x \in \mathrm{supp}(\boldsymbol{m}_X). \tag{7.17}$$

Thus there exists $n(N, D, \varepsilon)$ which does not depend on $X \in \mathcal{X}(N, D)$, such that we can find at most $n(N, D, \varepsilon)$ disjoint balls of radius ε in X. Thus $\mathrm{supp}(\boldsymbol{m}_X)$ can be covered by at most $n(N, D, \varepsilon)$ balls of radius 2ε. This means that the family $\mathcal{X}(N, D)$ is uniformly totally bounded, and thus it is compact w.r.t. Gromov-Hausdorff convergence (see e.g. Theorem 7.4.5 of [20]).

Pick a sequence $(X_n, d_n, \boldsymbol{m}_n) \in \mathcal{X}(N, D)$. By what we just proved, up to pass to a subsequence, not relabeled, we may assume that $(\mathrm{supp}(\boldsymbol{m}_n), d_n)$ converges in the Gromov-Hausdorff topology to some space (X, d). It is well known that in this situation there exists a compact space (Y, d_Y) and a family of isometric embeddings $f_n : \mathrm{supp}(\boldsymbol{m}_n) \to Y$, $f : X \to Y$, such that the Hausdorff distance between $f_n(\mathrm{supp}(\boldsymbol{m}_n))$ and $f(X)$ goes to 0 as $n \to \infty$.

The space $(f_n(\mathrm{supp}(\boldsymbol{m}_n), d_Y, (f_n)_\#\boldsymbol{m}_n))$ is isomorphic to $(X_n, d_n, \boldsymbol{m}_n)$ by construction for every $n \in \mathbb{N}$, and $(f(X), d_Y)$ is isometric to (X, d), so we identify these spaces with the respective subspaces of (Y, d_Y). Since (Y, d_Y) is compact, the sequence $(\boldsymbol{m}_n)$ admits a subsequence, not relabeled, which weakly converges to some $\boldsymbol{m} \in \mathscr{P}(Y)$. It is immediate to verify that actually $\boldsymbol{m} \in \mathscr{P}(X)$. Also, again by compactness, weak convergence is equivalent to convergence w.r.t. W_2, which means that there exists plans $\boldsymbol{\gamma}_n \in \mathscr{P}(Y^2)$ admissible for the couple $(\boldsymbol{m}, \boldsymbol{m}_n)$ such that

$$\int d_Y^2(x, \tilde{x})\, d\boldsymbol{\gamma}_n(x, \tilde{x}) \to 0.$$

Therefore $n \mapsto (d_Y, \boldsymbol{\gamma}_n)$ is a sequence of admissible couplings for $(X, d, \boldsymbol{m})$ and $(X_n, d_n, \boldsymbol{m}_n)$ whose cost tends to zero. This concludes the proof. □

Now we prove the HWI (which relates the entropy, often denoted by H, the Wasserstein distance W_2 and the Fisher information I) and the log-Sobolev inequalities. To this aim, we introduce the Fisher information functional $I : \mathscr{P}(X) \to [0, \infty]$ on a general metric measure space $(X, d, \boldsymbol{m})$ as the squared slope of the entropy $\mathscr{E}_\infty$:

$$I(\mu) := \begin{cases} \varlimsup\limits_{\nu \to \mu} \dfrac{\left((\mathscr{E}_\infty(\mu) - \mathscr{E}_\infty(\nu))^+\right)^2}{W_2^2(\mu, \nu)}, & \text{if } \mathscr{E}_\infty(\mu) < \infty, \\ +\infty, & \text{otherwise.} \end{cases}$$

The functional I is called Fisher information because its value on $(\mathbb{R}^d, |\cdot - \cdot|, \mathcal{L}^d)$ is given by

$$I(\rho\mathcal{L}^d) = \int \frac{|\nabla\rho|^2}{\rho} d\mathcal{L}^d,$$

and the object on the right hand side is called Fisher information on $\mathbb{R}^d$. It is possible to prove that a formula like the above one is writable and true on general $CD(K,\infty)$ spaces (see [7]), but we won't discuss this topic.

Proposition 7.18 (HWI inequality) *Let $(X, d, \boldsymbol{m})$ be a metric measure space satisfying the condition $CD(K,\infty)$. Then*

$$\mathscr{E}_\infty(\mu) \leq \mathscr{E}_\infty(\nu) + W_2(\mu,\nu)\sqrt{I(\mu)} - \frac{K}{2}W_2^2(\mu,\nu), \qquad \forall \mu,\nu \in \mathscr{P}(X). \tag{7.18}$$

In particular, choosing $\nu = \boldsymbol{m}$ it holds

$$\mathscr{E}_\infty(\mu) \leq W_2(\mu,\boldsymbol{m})\sqrt{I(\mu)} - \frac{K}{2}W_2^2(\mu,\boldsymbol{m}), \qquad \forall \mu \in \mathscr{P}(X). \tag{7.19}$$

Finally, if $K > 0$ the log-Sobolev inequality with constant K holds:

$$\mathscr{E}_\infty \leq \frac{I}{2K}. \tag{7.20}$$

Proof Clearly to prove (7.18) it is sufficient to deal with the case $\mathscr{E}_\infty(\nu), \mathscr{E}_\infty(\mu) < \infty$. Let (μ_t) be a constant speed geodesic from μ to ν such that

$$\mathscr{E}_\infty(\mu_t) \leq (1-t)\mathscr{E}_\infty(\mu) + t\mathscr{E}_\infty(\nu) - \frac{K}{2}t(1-t)W_2^2(\mu,\nu).$$

Then from $\sqrt{I(\mu)} \geq \overline{\lim}_{t\downarrow 0}(\mathscr{E}_\infty(\mu) - \mathscr{E}_\infty(\mu_t))/W_2(\mu,\mu_t)$ we get the thesis.

Equation (7.20) now follows from (7.19) and the trivial inequality

$$ab - \frac{1}{2}a^2 \leq \frac{1}{2}b^2,$$

valid for any $a, b \geq 0$. □

The log-Sobolev inequality is a notion of *global* Sobolev-type inequality, and it is known that it implies a global Poincaré inequality (we omit the proof of this fact). When working on metric measure spaces, however, it is often important to have at disposal a *local* Poincaré inequality (see e.g. the analysis done by Cheeger in [29]).

Our final goal is to show that in non-branching $CD(0,N)$ spaces a local Poincaré inequality holds. The importance of the non-branching assumption is due to the following lemma.

Lemma 7.19 *Let $(X, d, \boldsymbol{m})$ be a non branching $CD(0,N)$ space, $B \subset X$ a closed ball of positive measure and $2B$ the closed ball with same center and double radius. Define the measures $\mu := \boldsymbol{m}(B)^{-1}\boldsymbol{m}|_B$ and $\boldsymbol{\mu} := \gamma^{\cdot,\cdot}_\#(\mu \times \mu) \in \mathscr{P}(\mathrm{Geod}(X))$, where $(x,y) \mapsto \gamma^{x,y}$ is the map which associates to each x, y the unique geodesic connecting them (such a map is well defined for $\boldsymbol{m} \times \boldsymbol{m}$-a.e. x, y by Proposition 7.16). Then*

$$(\mathrm{e}_t)_\#\boldsymbol{\mu} \leq \frac{2^N}{\boldsymbol{m}(B)}\boldsymbol{m}|_{2B}, \qquad \forall t \in [0,1].$$

Proof Fix $x \in B$, $t \in (0,1)$ and consider the 'homothopy' map $B \ni y \mapsto Hom_t^x(y) := \gamma_t^{x,y}$. By Proposition 7.16 we know that this map is well defined for $\boldsymbol{m}$-a.e. y and that (using the characterization of geodesics given in Theorem 2.10) $t \mapsto \mu_t^x := (Hom_t^x)_\# \mu$ is the unique geodesic connecting δ_x to μ. We have

$$\mu_t^x(E) = \mu\big((Hom_t^x)^{-1}(E)\big) = \frac{\boldsymbol{m}\big((Hom_t^x)^{-1}(E)\big)}{\boldsymbol{m}(B)}, \qquad \forall E \subset X \text{ Borel.}$$

The non branching assumption ensures that Hom_t^x is invertible, therefore from the fact that $[\{x\}, (Hom_t^x)^{-1}(E)]_t = Hom_t^x\big(Hom_t^x)^{-1}(E)\big) = E$, the Brunn-Minkowski inequality and the fact that $\boldsymbol{m}(\{x\}) = 0$ we get

$$\boldsymbol{m}(E) \geq t^N \boldsymbol{m}\big((Hom_t^x)^{-1}(E)\big),$$

and therefore $\mu_t^x(E) \leq \frac{\boldsymbol{m}(E)}{t^N \boldsymbol{m}(B)}$. Given that E was arbitrary, we deduce

$$\mu_t^x \leq \frac{\boldsymbol{m}}{t^N \boldsymbol{m}(B)}. \tag{7.21}$$

Notice that the expression on the right hand side is independent on x.

Now pick $\boldsymbol{\mu}$ as in the hypothesis, and define $\mu_t := (\mathrm{e}_t)_\# \boldsymbol{\mu}$. The equalities

$$\int_X \varphi d\mu_t = \int_{\mathrm{Geod}(X)} \varphi(\gamma_t) d\boldsymbol{\mu}(\gamma) = \int_{X^2} \varphi(\gamma_t^{x,y}) d\mu(x) d\mu(y),$$
$$\int_X \varphi d\mu_t^x = \int_X \varphi(\gamma_t^{x,y}) d\mu(y),$$

valid for any $\varphi \in C_b(X)$, show that

$$\mu_t = \int \mu_t^x \, d\mu(x),$$

and therefore, by (7.21), we have

$$\mu_t \leq \frac{\boldsymbol{m}}{t^N \boldsymbol{m}(B)}.$$

All these arguments can be repeated symmetrically with $1-t$ in place of t (because the push forward of $\boldsymbol{\mu}$ via the map which takes γ and gives the geodesic $t \mapsto \gamma_{1-t}$, is $\boldsymbol{\mu}$ itself), thus we obtain

$$\mu_t \leq \min\left\{\frac{\boldsymbol{m}}{t^N \boldsymbol{m}(B)}, \frac{\boldsymbol{m}}{(1-t)^N \boldsymbol{m}(B)}\right\} \leq \frac{2^N \boldsymbol{m}}{\boldsymbol{m}(B)}, \qquad \forall t \in (0,1).$$

To conclude, it is sufficient to prove that μ_t is concentrated on $2B$ for all $t \in (0,1)$. But this is obvious, as μ_t is concentrated on $[B,B]_t$ and a geodesic whose endpoints lie on B cannot leave $2B$. □

As we said, we will use this lemma (together with the doubling property, which is a consequence of the Bishop-Gromov inequality) to prove a local Poincaré inequality. For simplicity, we stick to the case of Lipschitz functions and their local Lipschitz constant, although everything could be equivalently stated in terms of generic Borel functions and their upper gradients.

For $f : X \to \mathbb{R}$ Lipschitz, the local Lipschitz constant $|\nabla f| : X \to R$ is defined as

$$|\nabla f|(x) := \varlimsup_{y \to x} \frac{|f(x) - f(y)|}{d(x,y)}.$$

For any ball B such that $\boldsymbol{m}(B) > 0$, the number $\langle f \rangle_B$ is the average value of f on B:

$$\langle f \rangle_B := \frac{1}{\boldsymbol{m}(B)} \int_B f \, d\boldsymbol{m}.$$

Proposition 7.20 (Local Poincaré inequality) *Assume that $(X, d, \boldsymbol{m})$ is a non-branching $CD(0, N)$ space. Then for every ball B such that $\boldsymbol{m}(B) > 0$ and any Lipschitz function $f : X \to \mathbb{R}$ it holds*

$$\frac{1}{\boldsymbol{m}(B)} \int_B |f(x) - \langle f \rangle_B| \, d\boldsymbol{m}(x) \le r \frac{2^{2N+1}}{\boldsymbol{m}(2B)} \int_{2B} |\nabla f| d\boldsymbol{m},$$

r being the radius of B.

Proof Notice that

$$\begin{aligned} \frac{1}{\boldsymbol{m}(B)} \int_B |f(x) - \langle f \rangle_B| d\boldsymbol{m}(x) &\le \frac{1}{\boldsymbol{m}(B)^2} \int_{B \times B} |f(x) - f(y)| \, d\boldsymbol{m}(x) d\boldsymbol{m}(y) \\ &= \int_{\mathrm{Geod}(X)} |f(\gamma_0) - f(\gamma_1)| \, d\boldsymbol{\mu}(\gamma), \end{aligned}$$

where $\boldsymbol{\mu}$ is defined as in the statement of Lemma 7.19. Observe that for any geodesic γ, the map $t \mapsto f(\gamma_t)$ is Lipschitz and its derivative is bounded above by $d(\gamma_0, \gamma_1)|\nabla f|(\gamma_t)$ for a.e. t. Hence, since any geodesic γ whose endpoints are in B satisfies $d(\gamma_0, \gamma_1) \le 2r$, we have

$$\int_{\mathrm{Geod}(X)} |f(\gamma_0) - f(\gamma_1)| \, d\boldsymbol{\mu}(\gamma) \le 2r \int_0^1 \int_{\mathrm{Geod}(X)} |\nabla f|(\gamma_t) \, d\boldsymbol{\mu}(\gamma) dt = 2r \int_0^1 \int_X |\nabla f| d(\mathrm{e}_t)_\# \boldsymbol{\mu} dt.$$

By Lemma 7.19 we obtain

$$2r \int_0^1 \int_X |\nabla f| d(\mathrm{e}_t)_\# \boldsymbol{\mu} dt \le \frac{2^{N+1} r}{\boldsymbol{m}(B)} \int_{2B} |\nabla f| d\boldsymbol{m}.$$

By the Bishop-Gromov inequality we know that $\boldsymbol{m}(2B) \le 2^N \boldsymbol{m}(B)$ and thus

$$\frac{2^{N+1} r}{\boldsymbol{m}(B)} \int_{2B} |\nabla f| d\boldsymbol{m} \le \frac{2^{2N+1} r}{\boldsymbol{m}(2B)} \int_{2B} |\nabla f| d\boldsymbol{m},$$

which is the conclusion. □

7.3 Bibliographical notes

The content of this chapter is taken from the works of Lott and Villani on one side ([58], [57]) and of Sturm ([74], [75]) on the other.

The first link between K-geodesic convexity of the relative entropy functional in $(\mathscr{P}_2(M), W_2)$ and the bound from below on the Ricci curvature is has been given by Sturm and von Renesse in [76]. The works [74], [75] and [58] have been developed independently. The main difference between them is that Sturm provides the general definition of $CD(K, N)$ bound (which we didn't speak about, with the exception of the quick citation in Remark 7.9), while Lott and Villani focused on the cases $CD(K, \infty)$ and $CD(0, N)$. Apart from this, the works are strictly related and the differences are mostly on the technical side. We mention only one of these. In giving the definition of $CD(0, N)$ space we followed Sturm and asked only the functionals $\rho\boldsymbol{m} \mapsto N' \int (\rho - \rho^{1-1/N'}) d\boldsymbol{m}$,

$N' \geq N$, to be geodesically convex. Lott and Villani asked for something more restrictive, namely they introduced the *displacement convexity* classes DC_N as the set of functions $u : [0, \infty) \to \mathbb{R}$ continuous, convex and such that

$$z \qquad \mapsto \qquad z^N u(z^{-N}),$$

is convex. Notice that $u(z) := N'(z - z^{1-1/N'})$ belongs to DC_N. Then they say that a space is $CD(0, N)$ provided

$$\rho \boldsymbol{m} \qquad \mapsto \qquad \int u(\rho) d\boldsymbol{m},$$

(with the usual modifications for a measure which is not absolutely continuous) is geodesically convex for any $u \in DC_N$. This notion is still compatible with the Riemannian case and stable under convergence. The main advantage one has in working with this definition is the fact that for a $CD(0, N)$ space in this sense, for any couple of absolutely continuous measures there exists a geodesic connecting them which is made of absolutely continuous measures.

The distance $\mathbb{D}$ that we used to define the notion of convergence of metric measure spaces has been defined and studied by Sturm in [74]. This is not the only possible notion of convergence of metric measure spaces: Lott and Villani used a different one, see [58] or Chapter 27 of [80]. A good property of the distance $\mathbb{D}$ is that it pleasantly reminds the Wasserstein distance W_2: to some extent, the relation of $\mathbb{D}$ to W_2 is the same relation that there is between Gromov-Hausdorff distance and Hausdorff distance between compact subsets of a given metric space. A bad property is that it is not suitable to study convergence of metric measure spaces which are endowed with infinite reference measures (well, the definition can easily be adapted, but it would lead to a too strict notion of convergence - very much like the Gromov-Hausdorff distance, which is not used to discuss convergence of non compact metric spaces). The only notion of convergence of Polish spaces endowed with σ-finite measures that we are aware of, is the one discussed by Villani in Chapter 27 of [80] (Definition 27.30). It is interesting to remark that this notion of convergence does *not* guarantee uniqueness of the limit (which can be though of as a negative point of the theory), yet, bounds from below on the Ricci curvature are stable w.r.t. such convergence (which in turn is a positive point, as it tells that these bounds are 'even more stable')

The discussion on the local Poincaré inequality and on Lemma 7.19 is extracted from [57].

There is much more to say about the structure and the properties of spaces with Ricci curvature bounded below. This is an extremely fast evolving research area, and to give a complete discussion on the topic one would probably need a book nowadays. Two things are worth to be quickly mentioned.

The first one is the most important open problem on the subject: is the property of being a $CD(K, N)$ space a local notion? That is, suppose we have a metric measure space $(X, d, \boldsymbol{m})$ and a finite open cover $\{\Omega_i\}$ such that $(\Omega_i, d, \boldsymbol{m}(\Omega_i)^{-1}\boldsymbol{m}|_{\Omega_i})$ is a $CD(K, N)$ space for every i. Can we deduce that $(X, d, \boldsymbol{m})$ is a $CD(K, N)$ space as well? One would like the answer to be affirmative, as any notion of curvature should be local. For $K = 0$ or $N = \infty$, this is actually the case, at least under some technical assumptions. The general case is still open, and up to now we only know that the conjecture 30.34 in [80] is *false*, being disproved by Deng and Sturm in [32] (see also [11]).

The second, and final, thing we want to mention is the case of Finsler manifolds, which are differentiable manifolds endowed with a norm - possibly not coming from an inner product - on each tangent space, which varies smoothly with the base point. A simple example of Finsler manifolds is the space $(\mathbb{R}^d, \|\cdot\|)$, where $\|\cdot\|$ is any norm. It turns out that for any choice of the norm, the space $(\mathbb{R}^d, \|\cdot\|, \mathcal{L}^d)$ is a $CD(0, N)$ space. Various experts have different opinion about this fact: namely, there is no agreement on the community concerning whether one really wants or not Finsler geometries to be included in the class of spaces with Ricci curvature bounded below. In any case, it is interesting to know whether there exists a different, more restrictive, notion of Ricci curvature bound which rules out the Finsler case. Progresses in this direction have been made in [8], where the notion of spaces with *Riemannian Ricci* bounded below is introduced: shortly said, these spaces are the subclass of $CD(K,N)$ spaces where the heat flow (studied in [45], [53], [7]) is linear.

References

[1] A. AGRACHEV AND P. LEE, *Optimal transportation under nonholonomic constraints*, Trans. Amer. Math. Soc., 361 (2009), pp. 6019–6047.

[2] G. ALBERTI, *On the structure of singular sets of convex functions*, Calc.Var. and Part.Diff.Eq., 2 (1994), pp. 17–27.

[3] G. ALBERTI AND L. AMBROSIO, *A geometrical approach to monotone functions in* $\mathbf{R}^n$, Math. Z., 230 (1999), pp. 259–316.

[4] L. AMBROSIO, *Lecture notes on optimal transport problem*, in Mathematical aspects of evolving interfaces, CIME summer school in Madeira (Pt), P. Colli and J. Rodrigues, eds., vol. 1812, Springer, 2003, pp. 1–52.

[5] L. AMBROSIO AND N. GIGLI, *Construction of the parallel transport in the Wasserstein space*, Methods Appl. Anal., 15 (2008), pp. 1–29.

[6] L. AMBROSIO, N. GIGLI, AND G. SAVARÉ, *Gradient flows in metric spaces and in the space of probability measures*, Lectures in Mathematics ETH Zürich, Birkhäuser Verlag, Basel, second ed., 2008.

[7] ——, *Calculus and heat flows in metric measure spaces with ricci curvature bounded below*, preprint, (2011).

[8] ——, *Spaces with riemannian ricci curvature bounded below*, preprint, (2011).

[9] L. AMBROSIO, B. KIRCHHEIM, AND A. PRATELLI, *Existence of optimal transport maps for crystalline norms*, Duke Mathematical Journal, 125 (2004), pp. 207–241.

[10] L. AMBROSIO AND S. RIGOT, *Optimal mass transportation in the Heisenberg group*, J. Funct. Anal., 208 (2004), pp. 261–301.

[11] K. BACHER AND K. T. STURM, *Localization and tensorization properties of the curvature-dimension condition for metric measure spaces*, J. Funct. Anal., 259 (2010), pp. 28–56.

[12] J.-D. BENAMOU AND Y. BRENIER, *A numerical method for the optimal time-continuous mass transport problem and related problems*, in Monge Ampère equation: applications to geometry and optimization (Deerfield Beach, FL, 1997), vol. 226 of Contemp. Math., Amer. Math. Soc., Providence, RI, 1999, pp. 1–11.

[13] P. BERNARD AND B. BUFFONI, *Optimal mass transportation and Mather theory*, J. Eur. Math. Soc. (JEMS), 9 (2007), pp. 85–127.

[14] M. BERNOT, V. CASELLES, AND J.-M. MOREL, *The structure of branched transportation networks*, Calc. Var. Partial Differential Equations, 32 (2008), pp. 279–317.

[15] S. BIANCHINI AND A. BRANCOLINI, *Estimates on path functionals over Wasserstein spaces*, SIAM J. Math. Anal., 42 (2010), pp. 1179–1217.

[16] A. BRANCOLINI, G. BUTTAZZO, AND F. SANTAMBROGIO, *Path functionals over Wasserstein spaces*, J. Eur. Math. Soc. (JEMS), 8 (2006), pp. 415–434.

[17] L. BRASCO, G. BUTTAZZO, AND F. SANTAMBROGIO, *A benamou-brenier approach to branched transport*, Accepted paper at SIAM J. of Math. Anal., (2010).

[18] Y. BRENIER, *Décomposition polaire et réarrangement monotone des champs de vecteurs*, C. R. Acad. Sci. Paris Sér. I Math., 305 (1987), pp. 805–808.

[19] ——, *Polar factorization and monotone rearrangement of vector-valued functions*, Comm. Pure Appl. Math., 44 (1991), pp. 375–417.

[20] D. BURAGO, Y. BURAGO, AND S. IVANOV, *A course in metric geometry*, vol. 33 of Graduate Studies in Mathematics, American Mathematical Society, Providence, RI, 2001.

[21] L. A. CAFFARELLI, *Boundary regularity of maps with convex potentials*, Comm. Pure Appl. Math., 45 (1992), pp. 1141–1151.

[22] ——, *The regularity of mappings with a convex potential*, J. Amer. Math. Soc., 5 (1992), pp. 99–104.

[23] ——, *Boundary regularity of maps with convex potentials. II*, Ann. of Math. (2), 144 (1996), pp. 453–496.

[24] L. A. CAFFARELLI, M. FELDMAN, AND R. J. MCCANN, *Constructing optimal maps for Monge's transport problem as a limit of strictly convex costs*, J. Amer. Math. Soc., 15 (2002), pp. 1–26 (electronic).

[25] L. CARAVENNA, *A proof of sudakov theorem with strictly convex norms*, Math. Z., to appear.

[26] J. A. CARRILLO, S. LISINI, G. SAVARÉ, AND D. SLEPCEV, *Nonlinear mobility continuity equations and generalized displacement convexity*, J. Funct. Anal., 258 (2010), pp. 1273–1309.

[27] T. CHAMPION AND L. DE PASCALE, *The Monge problem in $\mathbb{R}^d$*, Duke Math. J.

[28] ——, *The Monge problem for strictly convex norms in $\mathbb{R}^d$*, J. Eur. Math. Soc. (JEMS), 12 (2010), pp. 1355–1369.

[29] J. CHEEGER, *Differentiability of Lipschitz functions on metric measure spaces*, Geom. Funct. Anal., 9 (1999), pp. 428–517.

[30] D. CORDERO-ERAUSQUIN, B. NAZARET, AND C. VILLANI, *A mass-transportation approach to sharp Sobolev and Gagliardo-Nirenberg inequalities*, Adv. Math., 182 (2004), pp. 307–332.

[31] C. DELLACHERIE AND P.-A. MEYER, *Probabilities and potential*, vol. 29 of North-Holland Mathematics Studies, North-Holland Publishing Co., Amsterdam, 1978.

[32] Q. DENG AND K. T. STURM, *Localization and tensorization properties of the curvature-dimension condition for metric measure spaces ii*, Submitted, (2010).

[33] J. DOLBEAULT, B. NAZARET, AND G. SAVARÉ, *On the Bakry-Emery criterion for linear diffusions and weighted porous media equations*, Comm. Math. Sci, 6 (2008), pp. 477–494.

[34] L. C. EVANS AND W. GANGBO, *Differential equations methods for the Monge-Kantorovich mass transfer problem*, Mem. Amer. Math. Soc., 137 (1999), pp. viii+66.

[35] A. FATHI AND A. FIGALLI, *Optimal transportation on non-compact manifolds*, Israel J. Math., 175 (2010), pp. 1–59.

[36] D. FEYEL AND A. S. ÜSTÜNEL, *Monge-Kantorovich measure transportation and Monge-Ampère equation on Wiener space*, Probab. Theory Related Fields, 128 (2004), pp. 347–385.

[37] A. FIGALLI AND N. GIGLI, *A new transportation distance between non-negative measures, with applications to gradients flows with Dirichlet boundary conditions*, J. Math. Pures Appl. (9), 94 (2010), pp. 107–130.

[38] A. FIGALLI, F. MAGGI, AND A. PRATELLI, *A mass transportation approach to quantitative isoperimetric inequalities*, Invent. Math., 182 (2010), pp. 167–211.

[39] A. FIGALLI AND L. RIFFORD, *Mass transportation on sub-Riemannian manifolds*, Geom. Funct. Anal., 20 (2010), pp. 124–159.

[40] N. FUSCO, F. MAGGI, AND A. PRATELLI, *The sharp quantitative isoperimetric inequality*, Ann. of Math. (2), 168 (2008), pp. 941–980.

[41] W. GANGBO, *The Monge mass transfer problem and its applications*, in Monge Ampère equation: applications to geometry and optimization (Deerfield Beach, FL, 1997), vol. 226 of Contemp. Math., Amer. Math. Soc., Providence, RI, 1999, pp. 79–104.

[42] W. GANGBO AND R. J. MCCANN, *The geometry of optimal transportation*, Acta Math., 177 (1996), pp. 113–161.

[43] N. GIGLI, *On the geometry of the space of probability measures in R^n endowed with the quadratic optimal transport distance*, 2008. Thesis (Ph.D.)–Scuola Normale Superiore.

[44] ——, *Second order calculus on $(\mathscr{P}_2(M), W_2)$*, Accepted by Memoirs of the AMS, 2009.

[45] ——, *On the heat flow on metric measure spaces: existence, uniqueness and stability*, Calc. Var. Partial Differential Equations, (2010).

[46] ——, *On the inverse implication of Brenier-McCann theorems and the structure of $\mathscr{P}_2(M), W_2)$*, accepted paper Meth. Appl. Anal., (2011).

[47] R. JORDAN, D. KINDERLEHRER, AND F. OTTO, *The variational formulation of the Fokker-Planck equation*, SIAM J. Math. Anal., 29 (1998), pp. 1–17 (electronic).

[48] N. JUILLET, *On displacement interpolation of measures involved in brenier's theorem*, accepted paper Proc. of the AMS, (2011).

[49] L. V. KANTOROVICH, *On an effective method of solving certain classes of extremal problems*, Dokl. Akad. Nauk. USSR, 28 (1940), pp. 212–215.

[50] ——, *On the translocation of masses*, Dokl. Akad. Nauk. USSR, 37 (1942), pp. 199–201. English translation in J. Math. Sci. 133, 4 (2006), 1381Ð1382.

[51] L. V. KANTOROVICH AND G. S. RUBINSHTEIN, *On a space of totally additive functions*, Vestn. Leningrad. Univ. 13, 7 (1958), pp. 52–59.

[52] M. KNOTT AND C. S. SMITH, *On the optimal mapping of distributions*, J. Optim. Theory Appl., 43 (1984), pp. 39–49.

[53] K. KUWADA, N. GIGLI, AND S.-I. OHTA, *Heat flow on alexandrov spaces*, preprint, (2010).

[54] S. LISINI, *Characterization of absolutely continuous curves in Wasserstein spaces*, Calc. Var. Partial Differential Equations, 28 (2007), pp. 85–120.

[55] G. LOEPER, *On the regularity of solutions of optimal transportation problems*, Acta Math., 202 (2009), pp. 241–283.

[56] J. LOTT, *Some geometric calculations on Wasserstein space*, Comm. Math. Phys., 277 (2008), pp. 423–437.

[57] J. LOTT AND C. VILLANI, *Weak curvature conditions and functional inequalities*, J. Funct. Anal., (2007), pp. 311–333.

[58] J. LOTT AND C. VILLANI, *Ricci curvature for metric-measure spaces via optimal transport*, Ann. of Math. (2), 169 (2009), pp. 903–991.

[59] X.-N. Ma, N. S. Trudinger, and X.-J. Wang, *Regularity of potential functions of the optimal transportation problem*, Arch. Ration. Mech. Anal., 177 (2005), pp. 151–183.

[60] F. Maddalena and S. Solimini, *Transport distances and irrigation models*, J. Convex Anal., 16 (2009), pp. 121–152.

[61] F. Maddalena, S. Solimini, and J.-M. Morel, *A variational model of irrigation patterns*, Interfaces Free Bound., 5 (2003), pp. 391–415.

[62] R. J. McCann, *A convexity theory for interacting gases and equilibrium crystals*, ProQuest LLC, Ann Arbor, MI, 1994. Thesis (Ph.D.)–Princeton University.

[63] R. J. McCann, *A convexity principle for interacting gases*, Adv. Math., 128 (1997), pp. 153–179.

[64] ——, *Polar factorization of maps on riemannian manifolds*, Geometric and Functional Analysis, 11 (2001), pp. 589–608.

[65] V. D. Milman and G. Schechtman, *Asymptotic theory of finite-dimensional normed spaces*, vol. 1200 of Lecture Notes in Mathematics, Springer-Verlag, Berlin, 1986. With an appendix by M. Gromov.

[66] G. Monge, *Mémoire sur la théorie des d'eblais et des remblais*, Histoire de lÕAcadémie Royale des Sciences de Paris, (1781), pp. 666–704.

[67] F. Otto, *The geometry of dissipative evolution equations: the porous medium equation*, Comm. Partial Differential Equations, 26 (2001), pp. 101–174.

[68] A. Pratelli, *On the equality between Monge's infimum and Kantorovich's minimum in optimal mass transportation*, Annales de l'Institut Henri Poincare (B) Probability and Statistics, 43 (2007), pp. 1–13.

[69] S. T. Rachev and L. Rüschendorf, *Mass transportation problems. Vol. I*, Probability and its Applications, Springer-Verlag, New York, 1998. Theory.

[70] R. T. Rockafellar, *Convex Analysis*, Princeton University Press, Princeton, 1970.

[71] L. Rüschendorf and S. T. Rachev, *A characterization of random variables with minimum L^2-distance*, J. Multivariate Anal., 32 (1990), pp. 48–54.

[72] G. Savaré, *Gradient flows and diffusion semigroups in metric spaces under lower curvature bounds*, C. R. Math. Acad. Sci. Paris, 345 (2007), pp. 151–154.

[73] G. Savaré, *Gradient flows and evolution variational inequalities in metric spaces*, In preparation, (2010).

[74] K.-T. Sturm, *On the geometry of metric measure spaces. I*, Acta Math., 196 (2006), pp. 65–131.

[75] ——, *On the geometry of metric measure spaces. II*, Acta Math., 196 (2006), pp. 133–177.

[76] K.-T. Sturm and M.-K. von Renesse, *Transport inequalities, gradient estimates, entropy, and Ricci curvature*, Comm. Pure Appl. Math., 58 (2005), pp. 923–940.

[77] V. N. Sudakov, *Geometric problems in the theory of infinite-dimensional probability distributions*, Proc. Steklov Inst. Math., (1979), pp. i–v, 1–178. Cover to cover translation of Trudy Mat. Inst. Steklov **141** (1976).

[78] N. S. Trudinger and X.-J. Wang, *On the Monge mass transfer problem*, Calc. Var. Partial Differential Equations, 13 (2001), pp. 19–31.

[79] C. Villani, *Topics in optimal transportation*, vol. 58 of Graduate Studies in Mathematics, American Mathematical Society, Providence, RI, 2003.

[80] ——, *Optimal transport, old and new*, Springer Verlag, 2008.

[81] Q. XIA, *Optimal paths related to transport problems*, Commun. Contemp. Math., 5 (2003), pp. 251–279.

[82] ——, *Interior regularity of optimal transport paths*, Calc. Var. Partial Differential Equations, 20 (2004), pp. 283–299.

[83] L. ZAJÍČEK, *On the differentiability of convex functions in finite and infinite dimensional spaces*, Czechoslovak Math. J., 29 (1979), pp. 340–348.